



# The Promises and Pitfalls of Genoeconomics

## Citation

Benjamin, Daniel J., David Cesarini, Christopher F. Chabris, Edward L. Glaeser, David I. Laibson, Vilmundur Guðnason, Tamara B. Harris, et al. 2012. The promises and pitfalls of genoeconomics. *Annual Review of Economics* 4:627-662.

## Published Version

doi:10.1146/annurev-economics-080511-110939

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:10137000>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# The Promises and Pitfalls of Genoeconomics

Daniel J. Benjamin<sup>a</sup>, David Cesarini<sup>b</sup>, Christopher F. Chabris<sup>c</sup>,  
Edward L. Glaeser<sup>d</sup>, David I. Laibson<sup>d</sup>

*Age, Gene/Environment Susceptibility-Reykjavik Study:*  
Vilmundur Guðnason<sup>e</sup>, Tamara B. Harris<sup>f</sup>, Lenore J. Launer<sup>f</sup>, Shaun Purcell<sup>g</sup>,  
and Albert Vernon Smith<sup>e</sup>

*Swedish Twin Registry:*  
Magnus Johannesson<sup>h</sup>, Patrik K.E. Magnusson<sup>i</sup>,

*Framingham Heart Study:*  
Jonathan P. Beauchamp<sup>j</sup> and Nicholas A. Christakis<sup>k</sup>

*Wisconsin Longitudinal Study:*  
Craig S. Atwood<sup>l</sup>, Benjamin Hebert<sup>m</sup>, Jeremy Freese<sup>n</sup>, Robert M. Hauser<sup>o</sup>, Taissa S. Hauser<sup>o</sup>

*Swedish Large Schizophrenia Study:*  
Alexander Grankvist<sup>i</sup>, Christina M. Hultman<sup>i</sup>, Paul Lichtenstein<sup>i</sup>

<sup>a</sup> Economics Department, Cornell University, and NBER. <sup>b</sup> Economics Department, New York University.

<sup>c</sup> Psychology Department, Union College. <sup>d</sup> Economics Department, Harvard University, and NBER. <sup>e</sup> Icelandic Heart Association. <sup>f</sup> NIH. <sup>g</sup> Center for Human Genetics Research, Massachusetts General Hospital. <sup>h</sup> Economics Department, Stockholm School of Economics. <sup>i</sup> Department of Medical Epidemiology and Biostatistics, Karolinska Institutet. <sup>j</sup> McKinsey Consulting. <sup>k</sup> Sociology Department, Harvard University, and Department of Health Care Policy, Harvard Medical School. <sup>l</sup> Department of Medicine, University of Wisconsin. <sup>m</sup> Economics Department, Harvard University. <sup>n</sup> Sociology Department, Northwestern University. <sup>o</sup> Sociology Department, University of Wisconsin-Madison.

This Draft: March 20, 2012

---

For helpful comments, we are grateful to Peter Visscher and attendees at the 2007 NIA Workshop on Refining Economic Phenotypes for Genetic Analyses; 2009 AEA Meetings; 2009 MRRC Conference; 2009 Behavior Genetics Association Annual Meeting; IZA/Volkswagen Foundation Workshop: Genes, Brains, and the Labor Market; NSF Workshop on Genes, Cognition, and Social Behavior; Using GWAS to Explore Fundamental Questions about Aging in the HRS Sample: An Expert Meeting; Workshop to Explore a Social Science Genetic Association Consortium; 2010 and 2011 Integrating Genetics and Social Science Workshops; and seminar audiences at Caltech, University of California–San Diego, and Wharton. We are grateful to Jón Steinsson for advice in early stages of this work and Jon Torfi Jonasson, Loftur Guttormsson, and Helgi Skúli Kjartansson for advice in coding the Icelandic education variable. We thank Melissa Bickerman, Yeon Sik Cho, Cara Costich, Geoffrey Fisher, Julia Goorin, Ólafur Garðar Halldórsson, Sarina Kumar, Alice Lee, Logan Pritchard, Nathaniel Schorr, Abhishek Shah, and Kristina Tobio for excellent research assistance. We thank the NIA/NIH through grant P01AG005842-20S2. The Swedish Twin Registry is supported by the Swedish Department of Higher Education, the European Commission (grant QLG2-CT-2002-01254), the Swedish Research Council, and the Swedish Foundation for Strategic Research E-mail: db468@cornell.edu.

## **Abstract**

This paper reviews existing research at the intersection of genetics and economics, presents some new findings that illustrate the state of “genoeconomics” research, and surveys the prospects of this emerging field. Twin studies suggest that economic outcomes and preferences, once corrected for measurement error, appear to be about as heritable as many medical conditions and personality traits. Consistent with this pattern, we present new evidence on the heritability of permanent income and wealth. Turning to genetic association studies, we survey the main ways that direct measurement of genetic variation across individuals is likely to contribute to economics, and we outline the challenges that have slowed progress in making these contributions. The most urgent problem facing researchers in this field is that most existing efforts to find associations between genetic variation and economic behavior are based on samples that are too small to ensure adequate statistical power. This has led to many false positives in the literature. We suggest a number of possible strategies for improving power and remedying this problem: (i) pooling datasets; (ii) using statistical techniques that exploit the greater information content of many genes considered jointly; and (iii) focusing on economically-relevant traits that are most proximate to known biological mechanisms.

## 1. Introduction

With the sequencing of the human genome in 2001 (Venter et al., 2001; Lander et al., 2001), and with the continual, rapid, ongoing development of new technologies for measuring and analyzing the genome, the study of genetics has been transformed. Until recently, almost no information was available about genetic variation across individuals. Now most of the common genetic variation can be inexpensively measured.

These advances in genetics are, in turn, transforming medical research. Some diseases have been found to be linked to single genetic mutations in specific genes (for example, Huntington's disease and Fragile X syndrome), which can be assayed for the purpose of diagnosing the disease, predicting age of onset and/or severity, and better understanding how treatment response varies as a function of genetic characteristics. In the case of "complex" diseases or conditions, such as macular degeneration and obesity, new methods are beginning to identify the ensembles of genes that, along with environmental forces, account for individual differences. Unfortunately, each genetic variant identified in these studies of complex traits typically explains only a small amount of variation in the trait, and therefore the genetic risk factors identified so far are insufficient for the purpose of accurate medical diagnosis. Instead, the main benefit comes from the identification of new biological pathways and targets for therapeutic intervention. In short, genetics research has identified "new biology" for many major diseases, including diabetes, cancer, schizophrenia, and others.

Social scientists—including psychologists, anthropologists, political scientists, and increasingly, sociologists and economists—have begun to measure genetic variation and study how it relates to individual behaviors and outcomes. The early work involved measuring just a few "candidate genes" in small samples of laboratory participants. Costs of genotyping have now fallen to the point where very comprehensive information on a person's genetic constitution can be obtained at a moderate cost. Consequently, some large-scale social science surveys, such as the Health and Retirement Study, are gathering such data, and others will likely do so soon. With these new data sources, the scale of research at the intersection of social science and genetics will surely explode.

The purpose of this paper is to review the research that has been done to date at the intersection of genetics and economics, or "genoeconomics" (Benjamin et al. 2007), to present

some new findings that illustrate the current state of genoeconomics research, and to survey the field's prospects.

In Section 1, we begin by developing a simple conceptual framework that defines some key terms and makes explicit some critical assumptions. In Section 2, we review the economic research conducted in the tradition of classical “behavior genetics”—primarily involving comparisons between identical and fraternal twins—that seeks to estimate “heritability” for economic measures: the fraction of the variance that can be explained by genetic factors. One of the remarkable implications from this work is that in modern Western societies, for most outcomes in life, over half of the resemblance of two biological siblings reared in the same family stems from their genetic similarity. Another main implication is that, despite arguably being more complex and “downstream” from biochemical variation than psychological traits like cognitive ability and personality that are the traditional realm of behavior genetics, economic outcomes and preferences appear to be about as heritable as those traits, once adjustment is made for measurement error (Beauchamp et al., 2011a; Beauchamp et al., 2011b).

In Section 3, we overview what we see as the four ways that the intersection of molecular genetics and economics promises ultimately to contribute to economics: (i) identifying and measuring latent traits, (ii) identifying biological mechanisms that influence economic behavior, (iii) providing exogenous measures of preferences and abilities that may be used as control variables or—more problematically—as instrumental variables, and (iv) predicting differential effects of policies across individuals with different genetic constitutions. We review the small but growing body of work to date that reports associations between specific genes and economic traits. We end the section by outlining the main challenges obstructing progress in genoeconomics and discuss different ways of confronting these challenges.

In Section 4, we illustrate some of these themes using examples from our own work. Using an Icelandic sample, the Age, Gene/Environment Susceptibility-Reykjavik Study, we searched for associations between a set of outcomes of interest to economists and a set of candidate genes previously associated with cognitive functions or known to be involved in the brain's decision-making circuitry. We found a promising association between a particular genetic variant and educational attainment. The association was biologically plausible, associated with cognitive function, and replicated in a non-overlapping sample from the same respondent population. The association then failed to replicate in three other samples. We further illustrate

the widespread *non*-replicability of candidate gene associations by reviewing a systematic study we conducted of previously published associations between cognitive ability and 12 candidate genes. Across three new, large samples, we are unable to replicate these associations. We close Section 4 by proposing a number of strategies for surmounting the challenges that face genoeconomic research. If the genoeconomics enterprise is to bear fruit, it is important that social scientists recognize the many methodological lessons that have been learned in medical genetics over the past decade regarding the frequency of false positives in genetic associations.

This paper extends the analysis of Benjamin et al. (2007) and Beauchamp et al. (2011a). Benjamin et al. (2007) provide an initial definition of genoeconomics and survey the potential contributions of genetic studies in economics at a time when no such studies had yet been performed. Beauchamp et al. (2011a) report results from a large-scale genetic association study of educational attainment, which failed to identify any replicable associations. Using those results as a case study, Beauchamp et al. (2011a) reach similar conclusions as the present paper regarding the inferential challenges in genoeconomic research. While this paper primarily reviews published research at the intersection of genetics and economics and offers our perspective on the emerging field, it also presents several new findings. Online Appendix I provides details regarding our new behavior genetic results (in Section 2), which use Swedish Twin Registry data to estimate the heritabilities of permanent income and wealth. Online Appendix II provides details regarding our molecular-genetic analysis (in Section 4) from the Icelandic sample.

## **2. Conceptual Framework**

We adopt a conceptual framework that serves three purposes: it defines genetics terms that we will use throughout the paper, it makes explicit the assumptions typically made in empirical work, and it will be used later to clarify the link between behavior genetics and molecular genetics. We omit many biological nuances in order to focus on the concepts that are critical to understanding the field.

Human DNA is composed of a sequence of about 3 billion pairs of nucleotide molecules, each of which can be indexed by its location in the sequence.<sup>1</sup> This long sequence—the human genome—has subsequences called **genes**. Humans are believed to have 20,000–25,000 genes. Each gene provides the instructions that are used for building proteins. These proteins affect the structure and function of all cells in the body.

At the overwhelming majority of locations, there is virtually no variation in the nucleotides across individuals. The segments of DNA where individuals do differ are called **genetic polymorphisms** (from the Greek *poly*, meaning “many,” and *morphisms*, meaning “forms”). For simplicity, our discussion here focuses on the most common kind of genetic polymorphism, called a **single-nucleotide polymorphism (SNP)**. SNPs are locations in the DNA sequence where individuals differ from each other in terms of a single nucleotide. A single gene may contain hundreds of SNPs, and SNPs are also found in DNA regions that are not part of genes. We index SNPs by  $j$ , and we let  $J$  denote the total number of SNPs in the genome (currently it is believed that  $J \approx 16$  million). Conceptually, we can think of other kinds of genetic polymorphisms in the same way as SNPs, so focusing on SNPs is not misleading.<sup>2</sup>

At the vast majority of SNP locations, there are only two possible nucleotides that occur. The nucleotide of a SNP that is more common in the population is called the **major allele**, and the nucleotide that is less common is called the **minor allele**. At conception, each individual inherits half of his or her DNA from the mother and half from the father. For a given SNP, one allele is transmitted from each parent. The gene, and hence the protein it produces, is affected by the genetic material received from both parents, but it does not matter which material came from which parent. Therefore, for each SNP, there are three possibilities: an individual has 0 minor alleles, 1 minor allele, or 2 minor alleles. This number is called the individual’s **genotype**, and for individual  $i$  for SNP  $j$ , we denote its value by  $x_{ij}$ .

---

<sup>1</sup> The human genome is divided into 23 chromosomes. Each cell (aside from egg and sperm cells) includes two copies of each chromosome, one inherited from the mother and one from the father (except in the case of the Y chromosome, which is inherited by males only and comes entirely from the father).

<sup>2</sup> Other kinds of genetic polymorphisms include insertions or deletions from the DNA sequence and variable numbers of repetitions of a series of nucleotides.

Fix some outcome of interest, e.g., educational attainment, income, risk preferences, or body mass index. Let  $y_i$  denote the value of this outcome for individual  $i$ . The simplest model of genetic effects posits that  $y_i$  is determined according to

$$(1) \quad y_i = \mu + \sum_{j=1}^J \beta_j x_{ij} + \epsilon_i,$$

where  $\mu$  is the mean value of  $y_i$  in the population;  $\beta_j$  is the effect of SNP  $j$ ; and  $\epsilon_i$  is the effect of exogenous residual factors. Equation (1) embeds a variety of assumptions. For example, the restriction that the effect of a genotype is linear in the number of minor alleles is a simplifying assumption that can be, and often is, relaxed. Below, we discuss some other important extensions of (1).

$\beta_j$  should be understood as the treatment effect from an experiment in which one SNP (and nothing else in equation 1) is changed at conception. While such experiments are conducted on non-humans, in humans this treatment effect is a hypothetical construct. If  $\beta_j \neq 0$  for some  $j$ , then SNP  $j$  is called a “causal SNP.” As an example, it is now believed that there is a causal SNP in a gene called FTO on body weight (Frayling et al. 2007).<sup>3</sup> There are many ways in which FTO could affect body weight, e.g., by coding for a protein involved in metabolism, or by affecting food preferences. Identifying the correct mechanism(s) is an active area of research.

The residual term,  $\epsilon_i$ , is often called the “environmental effect,” but this terminology is imprecise and potentially misleading. Since the genotypic effects may operate through environmental channels,  $\epsilon_i$  should be interpreted as the component of environmental factors that are not endogenous to genetic endowment (Jencks, 1980). For example, *if* the mechanism through which the FTO SNP affects body weight is a preference for energy-rich foods that leads to increased caloric intake (as suggested by Cecil et al., 2008), then the component of caloric intake that is genetically induced is not part of  $\epsilon_i$ .

---

<sup>3</sup> In a study of nearly 40,000 Caucasians, Frayling et al. (2007) found that individuals with two minor alleles of a particular SNP weigh 3 kilograms more than individuals with two major alleles. This SNP may or may not be causal, since other unmeasured, correlated SNPs in or near FTO could be the causal SNPs.



Two important assumptions implicit in (1) are quite strong and are therefore relaxed in richer models. First, the genotypes,  $x_{ij}$  and  $x_{ij'}$ , for two different SNPs,  $j$  and  $j'$ , may interact in affecting the outcome. Second, a genotype  $x_{ij}$  may interact with factors in  $\epsilon_i$  in affecting the outcome. It is often claimed that such “gene-gene interaction” and “gene-environment interaction” effects matter for many outcomes. Indeed, since the treatment effect of a genotype is not a structural parameter, it will vary with some environmental conditions. For example, Rosenquist et al. (2012) report that the effect of the FTO SNP depends strongly on birth cohort.

Most of this paper is concerned with potential contributions that the field of genetics could make to the field of economics. We note here, however, a potential contribution that economics could make to genetics. The modeling tradition in economics could help in moving beyond the crude statistical framework outlined here toward more structural models. For example, a structural model of FTO might allow it to affect the marginal utilities of different foods, and possibly also the (production) function that maps caloric intake to body weight. Such a model would make predictions regarding how the treatment effect of the SNP would vary as a function of the prices and income of an individual, and it might predict compensatory behaviors, such as more exercise to try to reduce elevated body weight. The estimated model could be used to make predictions about the effects of changes in the economic environment. More generally, insights from economics about how environments can amplify or dampen genetic effects (depending, e.g., on the degree of substitutability or complementarity) may help geneticists more accurately model, identify, and understand genetic mechanisms.

## 2. Behavior Genetics and Economics

Behavior genetics is a field of research concerned with understanding how genetic endowments taken as a whole explain individual-level differences in outcomes. In terms of equation (1), individual  $i$ 's genetic endowment is defined as  $g_i \equiv \sum_{j=1}^J \beta_j x_{ij}$ . The field of behavior genetics pre-dates the availability of genotypic data, and its methods treat genetic endowments as latent variables whose effects are inferred indirectly by contrasting the similarity in outcomes of different pairs of relatives. Much research in behavior genetics focuses on estimating *heritability*, defined for a given outcome as the ratio of the population variance in

genetic endowment to the population variance in the outcome,  $\frac{\text{Var}(g_i)}{\text{Var}(y_i)}$ .<sup>4</sup> If genetic endowment  $g_i$  is independent of residual factors  $\epsilon_i$ , then heritability can equivalently be expressed as the population  $R^2$  for regression equation (1).

Over the years, there have been a number of misguided attempts to draw policy conclusions from heritability estimates. Goldberger (1979) clarified the key issues by pointing out that high heritability of an outcome does *not* imply that policy is impotent in affecting the outcome (see Manski, 2011, for a more recent discussion of these issues). High heritability means that existing, naturally-occurring variation in  $\epsilon_i$  does not explain much of the variation in  $y_i$ . It does not rule out that a policy could cause a large change in the outcome. In Goldberger's (1979) famous example, even if the heritability of eyesight were 100%, prescribing eyeglasses would still be a policy that passes the cost-benefit test. Conversely, the fact that an outcome has low heritability does not imply that it is especially susceptible to influence by policy.

Despite these important interpretational caveats, we believe that there are several reasons why economists may be interested in knowing the heritability of economic outcomes.<sup>5</sup> First, heritabilities of income, educational attainment, etc., are descriptive facts that constrain the set of theories regarding heterogeneity in preferences and abilities that can be considered plausible. For example, high heritability estimates are challenging for “blank-slate theories” of human nature, which have featured prominently in much social science work (Pinker, 2002).

Second, the pervasive finding of non-negligible heritabilities for economic outcomes confirms the common concern that unobserved genetic endowments are potentially a large source of confounding in attempts to estimate the effect of environmental variables on outcomes of interest, e.g., the effect of parental income on children's outcomes. In the language of econometrics, parental genotypes are omitted variables that influence the child's genotypes,

---

<sup>4</sup> If equation (1) is generalized to allow for a non-linear effect of genotype and/or gene-gene interactions, then it becomes necessary to distinguish “narrow-sense heritability” (essentially the  $R^2$  from the most predictive linear combination of the genotypes) from “broad-sense heritability” (the  $R^2$  of the genetic effects from the population regression, which includes their non-linear effects). In the simple framework of equation (1), these two concepts coincide.

<sup>5</sup> In plant and animal breeding, heritability is a key quantity because it measures the effect a breeder can have on the mean outcome in the next generation by selecting which animals to breed. In humans, using heritability to predict the next generation's outcomes based on the current generation's outcomes is far more tenuous because the reduced-form relationships described in equation (1) are likely to have changed from one generation to the next.

influence the child's behavior, and influence the child's environmental exposures (through the pathway of parental behavior).

Finally, because heritability can be interpreted as the population  $R^2$  for regression equation (1), it quantifies the degree to which an individual  $i$ 's outcome could be predicted if the  $\beta_j$ 's were known and the  $x_{ij}$ 's were observed (Visscher, Hill, & Wray, 2008). This will become an increasingly relevant upper bound as DNA information becomes more widely available and better estimates of  $\beta_j$ 's become possible. More immediately, a more heritable outcome may be a better target for efforts to discover particular SNPs that affect it because, all else equal, a more heritable outcome is likely to have more SNPs of larger effect.

The most common method for estimating heritability is the twin study. Twin studies exploit the fact that there are two types of twins: monozygotic (MZ) twins, who are essentially identical genetically,<sup>6</sup> and dizygotic (DZ) twins, whose genetic endowments are as correlated as those of ordinary siblings. The markedly higher resemblance that is often observed for MZ twins when compared to DZ twins on an outcome is therefore often interpreted as evidence that genetic endowment explains some of the variation in the trait.

Under some strong assumptions, data on the outcome for MZ and DZ twin pairs can be used to obtain a quantitative estimate of heritability. In terms of the conceptual framework described above, begin by assuming that an individual's genetic endowment  $g_i$  is independent of residual factors  $\epsilon_i$ . Since the two members of an MZ twin pair,  $m$  and  $m'$ , have identical genetic endowments, the covariance of their outcomes is given by

$$(2) \quad \text{Cov}(y_m, y_{m'}) = \text{Var}(g_m) + \text{Cov}(\epsilon_m, \epsilon_{m'}).$$

Denoting the two members of an DZ twin pair by  $d$  and  $d'$ , the covariance of their outcomes is given by

---

<sup>6</sup> Even MZ twins are not 100% genetically identical because of mutations. Moreover, there are ways in which even individuals with identical genomes at conception biochemically diverge over time. For example, the genome develops a set of external instructions—the epigenome—that regulates protein production. As a result of heterogeneous environmental exposures, identical twins will have different epigenomes.

$$(3) \quad \text{Cov}(y_d, y_{d'}) = \frac{1}{2} \text{Var}(g_d) + \text{Cov}(\epsilon_d, \epsilon_{d'}).$$

The claim that  $\text{Cov}(g_d, g_{d'}) = \frac{1}{2} \text{Var}(g_d)$  is not an immediate consequence of the fact that DZ twins share half their DNA on average (see Falconer & Mackay, 1996, Ch. 9, for the proof of the claim). The argument relies on the restriction in equation (1) that the genotypes affect the outcome linearly and additively, and requires the additional assumption that parents mate randomly (i.e., “assortative mating” on genetic endowments is ruled out).

If the distribution of genetic endowments and residual factors is the same both among MZ twins and among DZ twins as among the general population, then all three groups have the same population variance of genetic endowments,  $\text{Var}(g_m) = \text{Var}(g_d) = \text{Var}(g_i)$ , as well as the same population variance of outcomes,  $\text{Var}(y_m) = \text{Var}(y_d) = \text{Var}(y_i)$ .

The final key assumption is that,

$$(4) \quad \text{Cov}(\epsilon_m, \epsilon_{m'}) = \text{Cov}(\epsilon_d, \epsilon_{d'}) \equiv \text{Cov}_\epsilon.$$

Following Jencks (1980), this is how we interpret what is informally called the “equal environments assumption.” It requires that the residual factors co-vary equally for MZ twins as for DZ twins. Of the several strong assumptions in twin studies, this one has generated the most controversy, in part because it is rarely defined precisely, and it is easy to misinterpret. Clearly, MZ twins experience a more similar environment than DZ twins do: e.g., they are more similar in college completion and career interests, and because they look the same, they may evoke more similar reactions from others. The terminology “equal environments assumption” misleadingly suggests that this greater similarity of MZ twins’ environments violates assumption (4).

However, to the extent that this similarity in environment is caused by the similarity in genetic endowment, it is not a violation of (4). Instead, assumption (4) would be violated if, e.g., social interactions with an MZ twin generate higher covariance in residual shocks. For example, because he is genetically identical, an MZ twin may learn more about his own preferences from his co-twin’s experiences than a DZ twin does. Stenberg (2011) discusses the conceptual issues in interpreting the equal environments assumption and surveys some of the attempts to interrogate it empirically.

Now, dividing through equations (2) and (3) by the respective population variances,

$$(5) \quad \frac{\text{Cov}(y_m, y_{m'})}{\text{Var}(y_m)} = \frac{\text{Var}(g_m)}{\text{Var}(y_m)} + \frac{\text{Cov}(\epsilon_m, \epsilon_{m'})}{\text{Var}(y_m)}.$$

$$(6) \quad \frac{\text{Cov}(y_d, y_{d'})}{\text{Var}(y_d)} = \frac{1}{2} \frac{\text{Var}(g_d)}{\text{Var}(y_d)} + \frac{\text{Cov}(\epsilon_d, \epsilon_{d'})}{\text{Var}(y_d)}.$$

Since  $\frac{\text{Cov}(y_m, y_{m'})}{\text{Var}(y_m)}$  and  $\frac{\text{Cov}(y_d, y_{d'})}{\text{Var}(y_d)}$ , the correlation in outcomes across MZ pairs and DZ pairs, can be estimated from a sample of twins, equations (2) and (3) define two moment conditions that jointly identify heritability. While more sophisticated estimation methods are available, the simplest heritability estimator is just to “double the difference” between the correlations,

$$(7) \quad \frac{\widehat{\text{Var}}(g_i)}{\widehat{\text{Var}}(y_i)} = 2 \left( \frac{\widehat{\text{Cov}}(y_m, y_{m'})}{\widehat{\text{Var}}(y_m)} - \frac{\widehat{\text{Cov}}(y_d, y_{d'})}{\widehat{\text{Var}}(y_d)} \right).$$

Additional moments can be computed from datasets with more sibling types, thereby allowing the identification of more realistic models which relax the equal-environment assumption, the assumption that the effects of genotypes are purely additive, that mating is random, and that genetic endowments do not interact with the environment.<sup>7</sup> For an illustration of some of these ideas in the context of income heritability, see Björklund, Jäntti and Solon (2005).

The moment conditions (5) and (6) also identify  $\frac{\text{Cov}_\epsilon}{\text{Var}(y_i)}$ , which is typically called the “common environmental component.” This is the proportion of population variance in the outcome explained by residual factors shared among twins. It is often interpreted as the proportion of population variance in the outcome explained by residual factors shared among siblings in general, an interpretation that requires the additional assumption that  $\text{Cov}_\epsilon$  is also the

---

<sup>7</sup> If equation (7) is used as the estimator, then positive assortative mating on genetic endowments will generate a downward bias in the estimate of heritability because such mating increases the covariance of the genetic endowments of DZ twins. By the same token, the presence of non-linear or non-additive effects of the genotypes on the outcome will cause an upward bias in the estimate of heritability because these effects decrease the covariance.

covariance in residual factors among non-twin siblings. Although viewed by geneticists as a byproduct of the twin method of estimating heritability, this common environmental component is of interest to economists: It is a descriptive statistic measuring how much existing variation in family rearing environments accounts for variation in outcomes.

A simple example helps build some intuition for why this variance partitioning will often imply non-negligible heritabilities for outcomes like income that are many steps removed in the chain of causation from genes and protein production (Jencks, 1980). Consider a large sample of identical twins who are separated at birth and then randomly assigned to families. Under these conditions, and if non-genetic shared experiences in the uterine environments are not a source of greater MZ similarity, then any resemblance between the two twins must ultimately be due to similarity in genetic endowment. In this case,  $\text{Cov}(\epsilon_m, \epsilon_{m'}) = 0$ , so heritability could be estimated merely by computing the correlation in outcomes. This example illustrates that heritability estimates capture not only “direct” genetic effects but also “indirect” effects that operate through environmental exposure that are endogenous to genetic endowments. For example, genotype may be a source of selection into environments which in turn affect outcomes; e.g., genetic variation in reading ability may be mediated by self-initiated exposure to books (Lee, 2010), which is ultimately caused by genetic influences on preferences or abilities. As another example, an individual’s genotype may evoke environmental responses, such as parental investments (Becker, 1993; Becker & Tomes, 1976; Lizzeri & Siniscalchi, 2008).

Taubman (1976) introduced twin studies into economics. In a sample of about 2,500 white male twins who were all army veterans, he estimated the heritability of income to be between 18% and 41%. The basic finding that income is moderately heritable has now been repeatedly replicated in a variety of samples, including non-twin samples (Rowe, Vesterdal & Rodgers, 1999, Björklund, Jäntti & Solon, 2005). Sacerdote (2010) provides a recent review of behavior genetic work in economics, including research on adoptees.

A string of recent papers have shown that measures of economic preferences, usually elicited from either incentivized experiments or surveys, have heritabilities in the 20–30% range (Wallace et al., 2007; Cesarini et al., 2008, 2009, 2010, forthcoming; Barnea et al., 2010), though two papers reported a considerably higher estimate (Zhong et al., 2009a; Zyphur et al., 2009). Differences in common environment explain little of the variation in these outcomes.

These estimates of heritability (and also those of common environment effects) are likely biased toward zero, however, because of measurement error. Evidence for this view comes from Beauchamp et al. (2011b), who analyze a dataset of responses from over 11,000 twins to a battery of survey questions on risk attitudes. A subset of the respondents answered the survey twice. Beauchamp et al. (2011b) find that after adjustment for measurement error (assessed through the subset of repeat respondents), heritability for various survey-based measures of risk-taking is estimated to lie in the neighborhood of 40–50%, quite similar to the consensus estimates for personality and intelligence (Bouchard & McGue, 2003; Jang et al., 1996). Just as the original estimates of parent-child correlations in income (Becker & Tomes, 1986) were later shown to be greatly attenuated by measurement error (Mazumder, 2005; Solon, 1992; Zimmerman, 1992), so it would appear that twin-based estimates of heritability that fail to adjust for measurement error are quite severely downward biased.

*Heritability of permanent income.* We now turn to an illustration of the use of twin study methods by reporting some new estimates of two variables which are of central interest in economics: permanent income and net wealth. Past work has tended to focus on the heritability of current income, but for the purpose of describing inequality in standard of living, economists are typically more interested in consumption, or permanent income, than in transitory income. Largely due to data limitations, existing studies have focused on the heritability of income measured during a single year (Taubman, 1976) or up to three years (Björklund, Jäntti, & Solon, 2005). Here, we present heritability estimates of income averaged up to twenty years. For expositional convenience, we relegate a detailed variable and sample description to Online Appendix I and only sketch the details here.

We use a Swedish sample of twins from the Screening Across the Lifespan Twin study (SALT), augmented with a small number of individuals who answered a survey administered by the registry in 1973 (Q73). The SALT sample is described in Lichtenstein et al. (2002) and is comprised of twins born between 1926 and 1958. We use panel data on income from 1968–2005, drawn from administrative records. We restrict attention to individuals for whom we have complete income data for the twenty-year period and whose average yearly income exceeded 1000 SEK (approximately \$150).<sup>8</sup> Such individuals constitute 94% of the original sample. For

---

<sup>8</sup>We drop individuals with very low measured income because we believe such low numbers are especially likely to reflect reporting error or suggest that the individual in question had sources of income which were not known to the

further information about the sample and summary statistics, see Online Appendix I. We use the natural logarithm of income, and we residualize on a second-order age polynomial to account for income differences across birth cohorts.

In this sample, the estimated MZ correlations for single-year log-income are 0.41 for men and 0.27 for women, which is roughly comparable to existing estimates based on U.S. data (Taubman, 1976). The male MZ correlation in our sample is a little higher than was the figure reported in Björklund, Jäntti, and Solon (2005). However, when we average over a longer time period, we find that both the MZ and DZ correlations rise, suggesting a larger role for genetic factors in explaining the variation in permanent income. In male MZ twins, the correlation rises from 0.41 to 0.63 and in female MZ twins, the correlation rises from 0.27 to 0.48. The DZ twin correlations also rise, but not as dramatically.

In these data, applying the simple double-the-difference estimator (7) typically produces a *negative* estimate of the family environment.<sup>9</sup> We therefore instead proceed by imposing the restriction that the family environment component is zero and obtain a rough estimate of heritability by taking the average of the MZ correlation and twice the DZ correlation. This estimator suggests that heritability increases from 0.37 to 0.58 in men as we move from single-year income to a twenty-year average. The corresponding figures for women are 0.28 and 0.46. These findings suggest that permanent income is more heritable than single-year income. This conclusion partly seems to reflect the fact that measurement error and transitory shocks generate a downward bias in estimates of heritability (Mazumder, 2005; Solon, 1992; Zimmerman, 1992), consistent with our earlier conjecture that the heritability estimates of many other economic outcomes are downward biased.

These patterns of correlations illustrate Turkheimer's (2000) "three laws of behavior genetics," which are stylized facts that summarize the broad pattern of empirical findings in several decades of behavior genetics studies. The first law states that all behavioral outcomes are

---

tax authorities. The threshold of 1000 SEK is arbitrary, but the results do not vary substantively as we vary the threshold.

<sup>9</sup> Even if the assumptions underlying the variance decomposition (described above) hold exactly, a negative estimate could occur in a particular sample due to sampling variation and in that case should be interpreted as essentially an estimate of zero. In this case, since the number of twin pairs is rather large, a more likely explanation is that the assumptions are violated. For example, a negative common environment component estimate could be generated by a failure of the assumption of purely additive genetic effects, which would depress the genetic covariance between DZ twins, or by a failure of the equal environment assumption. It is also possible that the measurement errors are more highly correlated in MZ twins.



heritable. For comparison with our estimates of around .50 for permanent income, the heritability of personality traits and cognitive abilities is about .40 to .60 (Plomin, Owen, & McGuffin, 1994), and the heritability of height is about .80 (e.g., Silventoinen et al., 2003). Indeed, while Turkheimer’s first law is stated qualitatively, it could be made quantitative: of the hundreds of outcomes that have been analyzed to date, almost all have heritabilities estimated between .20 and .80 (see Plomin et al., 2008, for a review). The second law states that common family environment explains less variance than genes do, and the third law states that a substantial part of the variance in the outcome is left unexplained by the sum of genetic and common environment effects. Our results are consistent with these second and third laws, as well.

*Heritability of wealth.* To study wealth, we use data from the SALTY survey, which was recently administered by the Swedish Twin Registry. There are a total of 11,418 usable responses, but the wealth questions we study here were only administered to approximately 40% of the survey respondents; for further information and summary statistics, see Online Appendix I. Both because this sample size is far smaller and because wealth data are generally noisier than income data, our results on wealth are much less precise. Nonetheless, we report these results because, as far as we are aware, this is the first estimate of the heritability of wealth.

We use responses to a series of questions in which survey respondents are asked to indicate their assets in various categories, as well as their total debt. Since wealth results tend to be very sensitive to a few outliers with extreme values, we apply two transformations to the data. The first, which is frequently recommended for wealth data (see, e.g., Pence, 2006), is the hyperbolic sine transformation,  $\sinh^{-1}(x) = \ln(x + \sqrt{x^2 + 1})$ . This transformation is used to reduce the influence of extreme observations while—unlike the log transformation commonly used for other kinds of data—still allowing for negative values. As a robustness check, we also report results with the variable transformed to have a normal distribution. Formally, we first percentile-rank transform the net wealth variable and then take the inverse of the standard normal distribution of the ranking. This ensures that the resulting variable is standard normal.

Table 2 reports MZ and DZ correlations for wealth. The sibling correlations in wealth are quite low and estimated with less precision than the income correlations because only a subset of the SALTY respondents were asked about their assets and debt. Indeed, in the analyses separately by sex, there is even an instance of the male DZ correlation being higher than the MZ correlation, which we believe is likely to reflect sampling variation. When we pool for males and

females, however, we find that the correlations in MZ twins are significantly higher than the DZ correlations, implying heritability levels that range from about .20 to .40. Nonetheless, given the small sample and the imperfect measurement, we interpret these findings cautiously.

### **3. Molecular Genetics and Economics**

Molecular genetics is the field of research that studies the structure and function of DNA. Unlike behavior genetics, which draws indirect inferences regarding the effect of genetic endowments as a whole, molecular genetics involves directly measuring the genotypes for particular SNPs. Genoeconomics is an emerging field that incorporates such molecular genetic data into economic research.

#### *The Promise of Genoeconomics*

In our view, genoeconomics will ultimately make significant contributions to economics. We emphasize the word *ultimately* because—as will become clear below when we discuss the pitfalls of genoeconomics—there are many challenges to be overcome before these contributions can be realized. Nonetheless, it is the transformative promise of genoeconomics that makes us believe that, despite the challenges, the enterprise is worth pursuing. We anticipate that the eventual contributions will fall into four main categories.

*Direct measures of previously latent parameters.* First, measuring genotypes will advance empirical analysis by providing direct and exogenous measures of preferences and abilities. For example, as discussed above, an individual's FTO genotype may be a measure of preference for fatty foods. Preferences and abilities are key parameters in many models but currently must usually be treated as latent, unobserved variables. In principle (though not yet in practice), genetic methods could be used to identify such key structural parameters. These genotypes could then be used as control variables or as objects of primary interest.

*Biological mechanisms.* Second, social scientists will use genotypic data to learn about the biological mechanisms that underlie behaviors of interest. One possibility is that the genetic data bear on existing hypotheses. For example, experiments in which humans are exposed to the neuropeptide oxytocin suggest that oxytocin causes trusting behavior (Kosfeld et al., 2005). This leads naturally to the hypothesis that variation in the gene *OXTR*, which encodes the receptor for oxytocin, may be related to variation in trust-related behaviors. Unfortunately, the reported

association between genetic polymorphisms in *OXTR* and trusting behavior (Israel et al., 2009) has not replicated (Apicella et al., 2010). Nonetheless, using genetic data to explore existing hypotheses may bear fruit, and we review a number of efforts along these lines below in the context of reviewing candidate gene studies.

Even more intriguingly, analysis of the genetic data might suggest new hypotheses. In medicine, unexpected genetic associations with age-related macular degeneration and Crohn's disease have led to discoveries of new biological pathways for these diseases (Hirschhorn, 2009). While it is difficult to anticipate new hypotheses, we suspect they will arise in economics. We speculate that likely discoveries will involve the nature of preferences. While economists often study individual differences in terms of heterogeneity in "fundamental" preference parameters such as relative risk aversion, the (exponential) discount rate, and a weighting parameter for altruism, these primitive preferences do not (yet) rest on biological foundations—these categories were proposed by economic theorists before the modern age of empiricism. Identifying genetic differences that predict heterogeneity in behavior may provide an empirical basis for decomposing (or even rearranging) crude concepts like risk aversion and discounting into more primitive attributes with biological microfoundations.

*Genes as control variables and/or instrumental variables.* Third, social scientists may use genetic markers as control variables, thereby improving the power of standard economic analysis. By controlling for variation that would otherwise be absorbed in residuals, economists will be able to lower the standard errors associated with estimates of *non-genetic* parameters.

It is also possible that economists will be able to use genes as "instrumental variables" (IVs) to infer the causal effect of (non-genetic) factor  $X$  on (non-genetic) factor  $Y$  using observational data. For example, in epidemiology, this approach has been used to argue that greater alcohol consumption causes higher blood pressure, using as IVs SNPs in genes that code for proteins involved in alcohol metabolism (Chen et al., 2008; for reviews of the genetic IVs in epidemiology, see Davey Smith & Ebrahim, 2003, and Lawlor et al., 2008).

There are already a number of economics papers that use genes as IVs (Fletcher & Lehrer, 2009; Ding et al., 2009; Norton & Han, 2008; and von Hinke Kessler Scholder et al., 2010). For example, Fletcher and Lehrer (2009) study the effect of mental health ( $X$ ) on academic achievement ( $Y$ ). In effect, the idea is to use the fact that genotypes for SNPs affecting mental health are randomly assigned among siblings within a family as a natural experiment. As

usual with IV, the credibility of the analysis depends on whether the assumptions underlying IV estimation are satisfied; we return below to discuss the weak instruments problem and the exclusion restriction in the context of genetic IVs.

*Targeting interventions.* Finally, genetic information could eventually be useful for targeting social-scientific interventions, much like it is beginning to be useful for targeting medical interventions. For example, if dyslexia can eventually be predicted sufficiently well by genetic screening, parents with children who have dyslexia-susceptibility genes could be given the option of enrolling their kids in supplementary reading programs, years before a formal diagnosis of dyslexia (see Schumacher et al., 2007, for a review of genetic predictors of dyslexia). For adults, it is generally feasible and more accurate to measure realized preferences and abilities directly rather than relying on genetic predispositions, at least when there is no incentive to misrepresent one's type. For this reason, in the realm of economics, targeting interventions is most likely to take the form of parents obtaining genomic information about their children and then creating a developmental environment that is most likely to cultivate the children's preferences and abilities.

### *Estimating Genetic Effects*

All of these potential payoffs involve knowing the genetic effect on an outcome of one or more particular SNPs. Therefore, most work in genoeconomics to date has been focused on estimating genetic effects, and that is likely to remain true for the foreseeable future. We will discuss how genetic effects are estimated, and then we will turn to the pitfalls of genoeconomics, most of which involve challenges of estimation and causal inference.

A naïve approach would be simply to estimate equation (1),  $y_i = \mu + \sum_{j=1}^J \beta_j x_{ij} + \epsilon_i$ . Even if one could measure all  $J$  SNPs in the genome, however, this regression would fail the rank condition (unless one had more than 16 million subjects!). For that reason, it is standard instead to run  $K \ll J$  separate regression,

$$(8) \quad y_i = \mu + \beta_j x_{ij} + \epsilon_i,$$

for each of  $K$  SNPs that have been measured in the sample.

If the genotypes,  $x_{i1}, x_{i2}, \dots, x_{ij}$ , were mutually uncorrelated and uncorrelated with  $\epsilon_i$ , then estimating equation (8) rather than the population regression equation (1) would nonetheless yield unbiased estimates of the genetic effect  $\beta_j$ . In fact, however, due to how DNA is transmitted from parents to child, the genotypes of SNPs physically close to each other on the genome are correlated, often very highly so. Consequently, a robustly non-zero  $\hat{\beta}_j$  estimated from equation (8) does not necessarily imply that the true  $\beta_j$  from equation (1) is non-zero. SNP  $j$  could be proxying for a nearby, correlated SNP—possibly a SNP that is not included among the  $K$  SNPs that have been measured in the sample. For this reason, finding a robust association is the first step in a longer process (not discussed here) of obtaining high-resolution data on the associated SNP and adjacent SNPs in order to identify which is the causal SNP.

The estimated coefficient on SNP  $j$  could also be biased if genotype  $x_{ij}$  is correlated with residual factors  $\epsilon_i$ . Dealing with this possible confound is an important practical issue that we discuss below (under “population stratification” in the discussion of pitfalls).

The two main research strategies when testing for genetic association, the candidate gene approach and the genome-wide association study (GWAS), correspond to the two ways that researchers choose *which*  $K$  SNPs to study.

### *The Candidate Gene Approach*

In a ***candidate gene*** study, a researcher specifies *ex ante* hypotheses about a small set of  $K$  SNPs (with  $K$  typically in the 1 to 30 range), runs regression equation (8) for each, and tests each of the null hypotheses that  $\beta_j = 0$ , usually at the conventional  $\alpha = .05$  significance level. Ideally, these hypotheses are derived from the known biological function of the SNP. In practice, the hypotheses are often based on previously-reported associations with the same outcome or a related outcome, or the choice of SNPs is a result of their availability in the dataset the researchers are using.

The candidate gene approach, or hypothesis-based approach, was the main research strategy in medical genetics prior to the availability of dense SNP chips that made it possible and relatively inexpensive to measure hundreds of thousands, or millions, of SNPs. Candidate gene studies still predominate in the social science literature. Most of the major early successes in medical genetics were candidate gene studies. For example, because the plaques found in the

brain of Alzheimer's disease patients contain apolipoproteins, researchers examined whether genotypes in the *APOE* gene, which codes for an apolipoprotein, are associated with Alzheimer's disease. These genotypes, based on combinations of two SNPs, are now the strongest known genetic predictors of Alzheimer's disease that are common polymorphisms, as opposed to rare mutations (Strittmatter et al., 1993; St. George-Hyslop, 2000).

While the hypothesis-based approach seems intuitively reasonable, aside from the minority of cases (like *APOE*) where the hypotheses are very direct, it has a poor track record in medical genetics. It is now widely accepted that candidate gene study findings typically fail to replicate. To take an example which seems typical of the general pattern, a recent study used a sample with more than 20,000 individuals to examine previously-reported genetic associations with lung function. Out of over 100 genes examined, only one published association was shown to be robust (Obeidat et al., 2011).

At least three factors seem to account for the apparently high rate of false positives produced by these studies. First, the sample sizes were often relatively small, and thus the statistical power low, in the studies that initially reported positive findings, as discussed further below. Second, when the hypothesis-based approach is applied to complex diseases (or human behaviors), the basis for the hypothesis is almost always less precise than a direct link between a disease- or trait-relevant protein and the gene that codes for it. Ten years ago, those hypotheses often seemed convincing nonetheless, but today they seem much less so with the benefit of hindsight. That is partly because there are now many more known SNPs that could be hypothesized *ex ante* to be relevant and partly because it has become clear that—*ex post*, once an association has been found—it is possible to come up with seemingly-plausible hypotheses about why almost any gene should be associated with the outcome of interest. And even if a plausible mechanism linking a *gene* to an outcome is identified, there is no guarantee that a particular SNP in the gene that is selected as a candidate will affect the gene's function in the necessary way. Third, publication bias—the tendency for positive findings, as opposed to non-findings, to be selectively reported by researchers and selectively published by journals—is magnified in genetic association research because the typical dataset has data on many outcomes and many SNPs. Hence false positives arise due to multiple hypothesis testing. Investigation of gene-gene and gene-environment interaction effects, while in theory well-motivated, in practice exacerbates the multiple hypothesis testing problem (see, e.g., Duncan & Keller, 2011).

Recognizing these concerns, a leading field journal, *Behavior Genetics*, has recently adopted strict standards for publication of candidate gene studies (Hewitt, 2011). To be considered for publication, candidate gene studies must be well-powered and must account for all sources of multiple hypothesis testing, and any new finding must be accompanied by a replication. Today, the consensus view among genetics researchers is that the results from candidate gene studies are intriguing but should be interpreted with great caution.

### *The Genome-Wide Association Study (GWAS)*

A **genome-wide association study (GWAS)** is an atheoretical exercise that consists of looking for associations between the outcome and all of the SNPs measured on a dense SNP chip (usually  $K > 500,000$ , and now typically  $K \approx 2,500,000$ ), without any prior hypotheses. The researcher runs regression (8) for each of the  $K$  SNPs and tests each of the null hypotheses that  $\beta_j = 0$  at the **genome-wide significance level**, which is  $\alpha = 5 \times 10^{-8}$ .

The correlation structure of SNPs in the human genome is now well-understood, and the GWAS approach exploits this understanding in two ways. First, the SNPs that are measured on a dense SNP chip are selected such that, jointly, they cover, or “tag,” much of the non-rare genotypic variation across SNPs in the genome. Second, while the human genome contains about 16 million SNPs, due to the correlation structure, there are “only” effectively about 1 million independent SNPs. The genome-wide significance threshold of  $5 \times 10^{-8}$  therefore approximates the appropriate Bonferroni-corrected significance threshold of  $.05/1,000,000$  (McCarthy et al., 2008).

GWAS have produced many of the recent major discoveries in medical genetics. For example, the *FTO* gene mentioned above had not been hypothesized to be linked to body weight, but it repeatedly turned up in GWAS results. As nothing was previously known about this gene, its codename was assigned to represent “**FaT** mass and **O**besity associated,” and intensive work has begun on discovering its biological functions (Tung & Yeo, 2011). To take another example, in type 2 diabetes, GWAS-derived genetic discoveries have implicated new biological mechanisms and linked the disease to other processes, such as circadian rhythms (see Billings & Florez, 2010).

To date, most published genoeconomics papers are candidate gene studies of some economic preference parameter or economic behavior measured in the laboratory. All but one (Apicella et al., 2010) of the studies focused on laboratory measures reviewed below are based on samples small than 500 subjects and in some cases smaller than 100 subjects. Ebstein et al. (2010) and Beauchamp et al. (2011a) also provide reviews of the work in genoeconomics to date.

The first genoeconomic association was reported by Eisenberg et al. (2007), who tested whether two genetic polymorphisms near dopamine receptor genes (*DRD2* and *DRD4*, respectively) were associated with performance on a hypothetical delay discounting task measuring time preferences. The polymorphism near *DRD2* had a significant association with estimated discount rates, and there was an interaction between the *DRD2* and *DRD4* polymorphisms (but no main effect of the *DRD4* polymorphism). Another early paper was Knafo et al. (2008), who were inspired by findings that genetic variation near the *AVPR1a* gene causes differences in the social behavior of voles (Hammock & Young, 2002; Hammock et al., 2005). In a sample of 203 university students, Knafo et al. found that dictator game giving was associated with variation in this gene. A number of genoeconomic papers quickly followed suit. These papers tend to study outcomes that can be classified into one of two broad categories: decision making under uncertainty or social preferences.

Several papers inspired by neuroimaging studies of decision making under risk looked for associations between genes involved in the regulation of the dopaminergic system and various measures of risk-taking. Kuhnen and Chiao (2009) and Dreber et al. (2009) independently reported an association between a particular polymorphism of the *DRD4* gene and behavior in incentivized laboratory measures of risk-taking. Neither Carpenter et al. (2011) nor Dreber et al. (forthcoming) replicate this reported association. Other papers, also motivated by neuroeconomic theories, have reported statistically significant associations between measures of risk-taking and candidate genes (Zhong et al., 2009b; Zhong et al., 2009c; Crisan et al., 2009; Roe et al., 2010; Frydman et al., 2011).

There have also been some reported associations with various measures of social preferences. Israel et al. (2009) report an association between a SNP in the gene *OXTR* and dictator game giving. Apicella et al. (2010) fail to replicate this result in a larger sample and discuss possible explanations for the failed replication. McDermott et al. (2009) designed an



experiment in which 78 genotyped subjects were told that their earnings from a vocabulary task had been reduced by an anonymous third party. Subjects were then offered the opportunity to punish the third party. The subjects were either told that 80% or 20% of their earnings had been taken by the third party. *MAOA* genotype predicted the behavioral response only following the more aggressive provocation. Finally, Zhong et al. (2010) report that an interaction between a *DRD4* polymorphism and season of birth affects responder behavior in the ultimatum game.

A handful of papers have examined associations between candidate genes and behaviors and outcomes outside the laboratory, such as credit card debt (De Neve et al., 2010), portfolio risk (Kuhnen et al., 2011), happiness (De Neve et al., 2011), and self-employment (Nicolaou et al., 2011). In a large sample, van der Loos et al. (2011) fail to replicate the reported association with self-employment.

Beauchamp et al. (2011a) is the only example of a GWAS published in an economics journal to date, although van der Loos et al. (2010) describe an ongoing study. In a GWAS of educational attainment with a sample of 7,574 Framingham Heart Study participants, Beauchamp et al. (2011a) report 20 associations that fell short of genome-wide significance. They also report a replication attempt with a sample of 9,535 individuals from a Dutch sample. None of the 20 SNP associations replicated at the .05 significance level, and only 9 of 20 even had the same sign. Martin et al. (2011) report on the results for a GWAS of educational attainment in a sample of 9,538 Australians and also fail to find any genome-wide significant associations.

### *The Pitfalls of Genoeconomics*

Despite the recent explosion in the number of papers reporting genotype-behavior associations, we are pessimistic about the replicability of most findings to date. The most urgent problem—discussed in detail below—is that the most persuasive evidence suggests that true genotype-behavior associations have tiny effect sizes, so current research designs in the social sciences are woefully underpowered. However, even once this problem has been solved, there are a number of further obstacles that must be overcome before the promises of genoeconomics mentioned above—providing direct measures of latent parameters, elucidating biological mechanisms, using genes as controls or IVs, and targeting interventions—can be realized.

*Causal inference.* The biological-mechanisms and genes-as-IVs promises require uncovering the causal effect of particular SNPs on behavior, but most existing research designs focus on detecting correlations. There are myriad confounds to a causal interpretation. As discussed above, because of the way DNA is transmitted from parents to children, the genotype of a SNP is often highly correlated with the genotypes of nearby SNPs, necessitating follow-up work to any robustly-detected association in order to identify which SNP is actually responsible. Another common confound is that an individual's genotype is correlated with his parents' genotypes, which are in turn correlated with the individual's family environment. For example, a SNP may be associated with cognitive ability even though it actually causes nurturing behavior; an individual with the nurturing genotype is likely to have parents with that genotype, whose bias toward nurturing behavior may lead them to create a family environment that potentiates the development of higher cognitive ability.

In practice, the most common concern is confounding from ***population stratification***: different groups within the sample differ in allele frequencies and also differ in their outcome for non-genetic reasons. A famous pedagogical example is the “chopsticks effect” (Lander & Schork, 1994): a study concerned with finding the genetic causes of chopstick use would find a significant association for any SNP whose allele frequencies differs appreciably between Asians and non-Asians, even though most variation in chopstick use is explained by cultural factors. This example might seem to suggest that a simple fix would be to control for race or ethnicity. Indeed, it is standard practice to restrict a genetic association study to subjects of a common ethnic background. It has been found, however, that allele frequencies can differ even within ethnically homogeneous populations, such as different regions within Iceland (Price et al., 2009). For this reason, it is a common practice in GWASs to include as control variables the first 10 or 20 principal components of all the genotypes measured in the dense SNP chip. These principal components seem to pick up much of the subtle genetic structure within a population (Price et al., 2006). A disadvantage of candidate gene studies relative to GWAS designs is that they often do not have dense SNP data and hence cannot control for subtle genetic differentiation using principal components.

In our view, building the case that a robustly-identified association is causal will take time and will require convergent evidence from various research strategies. For ruling out a number of potential confounds, it would be useful to have evidence for a genetic association in a

dataset that includes siblings, using regression equation (8) but with family fixed effects. When identifying off of within-family variation, population stratification ceases to be a concern. Moreover, genotypes are randomly assigned to siblings who share the same biological parents. Complementary with such empirical evidence would be experimental evidence from animal models, where genotypes can be experimentally modified at conception, as well as biological evidence on the function of protein products of the gene.

*Pleiotropy.* There is a further obstacle to credibly using genes as IVs. For the exclusion restriction to be satisfied, the causal effects of the genes must be understood well enough to rule out alternative pathways (besides  $X$ ) by which the genes could affect outcome  $Y$ . Since many genes code for proteins that have multiple functions and effects—a phenomenon called *pleiotropy* that in most cases biologists have barely begun to understand—it seems unlikely that we can be confident about all of the consequences of any particular genotype in the foreseeable future (Conley, 2009).

*Missing heritability.* Targeting interventions is one of the potential contributions closest at hand because the genetic markers can be merely predictive, rather than causal, and because an index composed of many SNPs can be used, which may in the aggregate have substantial predictive power even if any constituent SNP in the index has little or none.<sup>10</sup> However, while we expect eventual successes, it will likely be slow and challenging to find sufficient predictive power even from an index.

In medical genetics, with the exception of a few, rare, single-gene disorders, there has been a general failure to find sizeable aggregate predictive power from the associated genetic markers identified to date—a problem now called the “missing heritability” puzzle (see, e.g. Purcell et al., 2009). Consider height, a highly-studied physical trait that both is measured with much less error than behavioral traits and is more heritable. Behavioral genetics studies on twins and other relatives indicate that about 80% of the variability in height is due to genetic factors.

---

<sup>10</sup> The standard method of constructing a predictive index (Purcell et al., 2009) is to take the coefficients,  $\hat{\beta}_1, \dots, \hat{\beta}_K$ , estimated from running regression (8) for each of the  $K$  SNPs; keeping only a subset,  $Q$  of the coefficients with  $Q < K$  such that the genotypes of these  $Q$  SNPs are approximately uncorrelated; and then forming a predictor  $\hat{y}_i$  for each individual  $i$  using an analog of equation (2),  $\hat{y}_i = \hat{\mu} + \sum_{j=1}^Q \hat{\beta}_{q(j)} x_{i,q(j)}$ , where  $q(j)$  is the  $j$ th SNP in the subset of  $Q$  SNPs. There are different methods of choosing the subset  $Q$ , e.g., choosing only SNPs whose  $p$ -value from regression (8) is below some threshold. Predictive power is assessed as the  $R^2$  from a regression of  $y_i$  on  $\hat{y}_i$  in a new sample.

Furthermore, recent estimates suggest that, even just using the SNPs measured with current dense SNP genotyping technologies (which leave non-SNP genetic polymorphisms unmeasured), it should be possible to predict 45% of the variance in human height (Yang et al., 2010). Yet the aggregate predictive power from known genotypes is only about 10%, with 0.3% being the largest  $R^2$  of any one of the SNPs in 180 separate locations in the genome so far found to be associated with height (Lango Allen et al., 2010). This state of affairs for height, and similar states of affairs for a variety of intensively-studied medical outcomes, suggests that for these outcomes, the bulk of the genetic variance is carried by many SNPs of miniscule effects that are spread diffusely throughout the genome. If so, unrealistically large sample sizes may be required to identify all of these SNPs. Given the failure to find sizeable predictable power in physical traits, the challenge is likely to be at least as large for behavioral traits where the causal mechanisms are probably more complex.

*Low power.* The most urgent problem, however, is that most efforts in the social sciences to discover genetic associations are underpowered. The fundamental reason is that almost every *true* genotype-behavior correlation is probably very small. For example, consider cognitive ability. It is among the most reliably-measured and widely-studied outcomes in social-science genetics, yet it is unclear whether any purported genetic associations with cognitive ability are robust. In a meta-analysis of 67 independent samples, variation in the *COMT* gene was found to explain 0.1% of the variance in cognitive ability, although even this estimate is likely to be biased upward because the meta-analysis found evidence of publication bias (Barnett, Scoriels, & Munafò, 2008). Even if this effect size were correct, since the strongest associations are more likely to be discovered first, most of the SNPs truly associated with cognitive ability probably have smaller effects. As another example, a recent GWAS of the classic “Big Five” personality traits (neuroticism, extraversion, openness, agreeableness, and conscientiousness) with a sample size of about 20,000 individuals failed to find any genome-wide significant associations (de Moor et al., 2012).

To get a sense of the magnitude of the problem, consider a candidate gene study of a particular SNP. To simplify, suppose there are only two genotypes for the SNP, with carriers of the High variant, as opposed to carriers of the Low variant, hypothesized to have a higher value for the outcome. To further simplify, suppose there are only two possibilities: either there is a true association, or there is not. Imagine the outcome is distributed normally. Suppose it is

known that, if there is an association, then the SNP explains  $R^2 = 0.1\%$ —a rather large effect size for a single SNP (the same size as the *COMT* association with cognitive ability). A first question is: what sample size is required for the typical benchmark of 80% power to detect the effect using regression (8) at the standard, two-tailed .05 significance level? The answer: 7,845. This is far larger than typical samples to date in geno-economics, which have numbered from less than one hundred to several hundred in studies using laboratory measures and a few thousand in studies using non-laboratory data.

Now suppose that in a sample of size  $N$ , a researcher observes a statistically significant association at the standard .05 significance level. How large does  $N$  have to be in order for this result to constitute substantial evidence about whether there is an association? Panel A of Table 3 shows how a researcher's posterior belief (after having seen the data) that there is a true association should depend on the researcher's prior belief and on  $N$ . Of course, it is difficult to know what an appropriate prior belief is, but for a typical candidate SNP, it is probably much less than 10%. In a GWAS where millions of SNPs are tested, the prior probability that a typical given SNP has a true relationship is less than .01%.

A proper Bayesian thinker would barely update her posteriors at all when faced with a statistically significant association in a sample of 100 individuals. Because the effect size is so small, statistical power—the probability of finding a statistically significant association under the alternative hypothesis that there is truly a relationship—is only 6%. At a significance level of .05, there is a 5% probability of finding a statistically significant association under the null hypothesis. Hence finding a statistically significant association at the .05 level is almost equally likely under the null hypothesis as under the alternative hypothesis and hence is essentially uninformative regarding which hypothesis is more likely to be correct.

In a sample of 30,000 individuals, where statistical power is 99%, the likelihood of finding a statistically significant association under the alternative hypothesis is about twenty times the likelihood of finding a statistically significant association under the null hypothesis. When the prior probability of a true association is 0.01%, the posterior probability after observing a statistically significant association is 0.20%, which is unfortunately still extremely low. Even if the prior probability of a true association were as high as 10%, the posterior probability after observing a statistically significant association would be 69%, leaving a 31% chance that the reported association is a false positive.

Because the effect sizes are so small, these calculations defy our usual expectations about the robustness of statistically significant findings and suggest that, when evaluating candidate gene studies, it is valuable to conduct such calculations rather than rely on our faulty intuitions. The clear message from Panel A of Table 3 is that a researcher should conclude almost nothing about a genotype-behavior relationship from a sample size in the hundreds, and sample sizes must number in the many thousands before non-trivial inferences are appropriate.<sup>11</sup>

Relative to complex behavioral outcomes, the power challenge is less daunting for intermediate outcomes, such as functional Magnetic Resonance Imaging (fMRI) data, for which effects of individual SNPs are probably larger, but an adequately-powered candidate gene study still requires sample sizes much larger than is currently typical. For instance, suppose it is known that, if there is an association, then the SNP explains  $R^2 = 3\%$  (i.e., thirty times more variance than the above-estimated effect of *COMT* on cognitive ability). Under the same assumptions as above, for the conventional 80% power level, a sample size of  $N = 258$  is required. However, due to the cost of using MRI scanners, a typical large fMRI study currently has a sample size of about  $N = 100$ .

#### 4. Cautionary Tales and Constructive Responses

In this section, we illustrate some of the challenges of genoeconomics research by telling two cautionary tales that trace out the trajectory of our research projects in this area, and outline three constructive responses.

##### *An Icelandic Saga*

When we began our work on genoeconomics approximately ten years ago, before dense SNP chips became relatively inexpensive, the standard empirical strategy in the medical genetics literature was the candidate gene approach, so we followed the same methodology. At the time, there were extremely few datasets that contained both economic and genotypic data. No

---

<sup>11</sup> The power challenge is probably less daunting for functional Magnetic Resonance Imaging (fMRI) data, for which effects of individual SNPs are probably larger, but reasonable power still requires sample sizes much larger than is currently typical. For instance, suppose it is known that, if there is an association, then the SNP explains  $R^2 = 3\%$ . Under the same assumptions as above, a sample size of  $N = 258$  is required for 80% power.

economic datasets had collected genotypic data, but we were fortunate to team up with the Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-RS), an Icelandic medical study that happened to have collected several survey measures of interest to economists. Here, we sketch our analysis of this data; full details are available in Online Appendix II.

Constrained by what was available in the data, we constructed the following eight “economic outcomes” that serve as dependent variables in the analysis:

1. *Time preference index*: An index of present-oriented behaviors, combining measures of alcohol use, cigarette use, and body mass index (BMI) at age 25.
2. *Happiness*
3. *Self-reported health*
4. *Housing wealth*
5. *Human capital index*: An index of human capital, combining years of education with number of foreign languages learned.
6. *Income* (predicted by occupation held at mid-life)
7. *Labor supply*
8. *Social capital index*: An index of social capital, combining amount of regular contact with relatives and friends, attendance at religious services, and participation in social activities.

We then created a list of candidate genes that were most likely to be related to economic decision-making, given what was known at the time our study was initiated. We obtained enough funding to have blood samples from 2,349 AGES-RS participants run through a custom-designed microarray that could measure 384 SNPs. We chose which genes to study based on two criteria: published associations with cognition-related outcomes or disorders (e.g., cognitive ability, long-term memory, Alzheimer disease, schizophrenia, ADHD), and/or membership in the dopamine or serotonin neurotransmitter systems. If the gene was small enough, we included enough SNPs to capture most of the possible variation in that gene. If the gene was too large, we included only the SNPs on the gene that had been specifically mentioned in published association studies. We supplemented the 384 SNPs we specified with several additional SNPs that had been previously genotyped in AGES-RS for other purposes (e.g., the two SNPs in APOE that define the genotypes associated with late-onset Alzheimer’s disease). Adding these

additional SNPs, and subtracting the few SNPs that failed to genotype correctly, our total number of SNPs was 415 in a total of 68 genes.

We ran regression (8) 3320 times, one for each of the 8 outcome  $\times$  415 candidate SNP combinations. The three most statistically significant associations are: the “social capital index” with a SNP called rs17529477 in the gene *DRD2* ( $p < .0005$ ), the “time preference index” with rs908867 in *BDNF* ( $p < .0001$ ), and the “human capital index” with rs2267539 in *SSADH* ( $p < .001$ ). The results are virtually identical when linear controls for age and sex, the standard control variables in medical genetics, are included in the regressions.

Naturally the standard  $p$ -values (reported above) from such regressions are easily misinterpreted due to multiple hypothesis testing. In 2008, we were able to attempt to replicate these three “top hits” in a non-overlapping sample of 1,759 AGES-RS participants who had been genotyped using a dense SNP chip (the Illumina Hu370CNV) for a different research project. While that chip did not directly measure any of the three SNPs that exhibited a promising association, it is standard in genetics to impute data on missing SNPs using observed data on surrounding SNPs, which is usually highly accurate because of the high correlation among nearby SNPs. While the imputation quality for *DRD2* rs17529477 was relatively low, we were able to impute the other two SNPs with very high accuracy.<sup>12</sup>

The association between the time preference index and *BDNF* rs908867 did not replicate ( $p = .531$ ). The association between the human capital index and *SSADH* rs2267539 not only replicated ( $p = .02$ ), but has similar effect sizes in the two samples: a coefficient of .23 with standard error of .07 in the first sample, and a coefficient of .19 with standard error of .08 in the second sample. Combining the first and second samples, this association has an  $R^2$  of 0.47%, which is quite large for an individual SNP.

The left panel of Figure 1 shows, for the first and second samples combined, the average level of the “human capital index” by genotype. In this case, it turns out that the relationship

---

<sup>12</sup> One commonly-used metric for imputation quality is the “variance ratio”: The ratio of the variance across individuals in the imputed genotype to the expected binomial variance based on frequency of the minor allele. In a large sample, an accurate imputation will have a variance ratio of 1, while an imputation based on no information will have a variance ratio of 0. In standard GWAS sample sizes of several thousand individuals, most imputed SNPs have variance ratios above 0.9 because of the generally high degree of correlation with nearby SNPs. The variance ratios for *DRD2* rs17529477, *BDNF* rs908867, and *SSADH* rs2267539 were 0.657, 0.999, and 0.956, respectively. While a variance ratio is not generally considered unacceptably low unless it is below 0.3, we were suspicious about the *DRD2* SNP imputation because the concordance rate—the fraction of matches between imputed genotype and known genotype in the part of the GWAS sample that overlapped with the candidate gene sample—was only 81%.



between the level of the index and the number of A alleles is monotonic. To give a sense for the magnitude of the relationship in natural units, the mean years of education—the main constituent of the index—is given in parentheses: years of education for G/G participants were 8.3, and this increased to 8.8 for A/G participants and 8.9 for A/A participants. Table 6 shows, for the combined sample, regression specification (8) with controls for population stratification (the first two principal components of the dense SNP data) and regional variation in education.

An association is less likely to be a false positive if there is a plausible biological mechanism for the relationship. The gene SSADH (also known as ALDH5A1) codes for an enzyme that metabolizes GABA, the principal inhibitory neurotransmitter in the brain. This gene matters for cognition: it has been associated with general cognitive ability (IQ; Plomin et al., 2004), it is related to the preservation of cognitive function in the elderly (De Rango et al., 2008), and it may be undergoing recent natural selection (Leone et al., 2006), as might be expected for a gene that has a large effect on a trait that could assist in survival and reproduction. Furthermore, rare mutations of SSADH are associated with mental retardation, and animals in which the gene is experimentally “knocked out” (i.e., rendered inoperative) are cognitively impaired and develop epileptic seizures (Knerr et al., 2008; Buzzi et al., 2006).

If the gene is related to our human capital index via its effect on cognitive ability, then we should observe that cognitive ability mediates the relationship between the SNP and human capital. To directly test this mechanism, we constructed a measure of cognitive ability using a variety of cognitive tests that had been administered to AGES-RS participants (see Online Appendix II for details).

The right panel of Figure 1 shows that, as expected, this index of cognitive ability is associated with the SNP of interest, SSADH rs2267539. In a regression (8) with standardized cognitive ability as the outcome, the coefficient is .11 with a standard error of .03, indicating that a switch of one G allele to an A allele is associated with one-ninth of a standard deviation greater cognitive ability (corresponding to about 1.7 points on the IQ scale). This association is highly statistically significant, with a relatively large  $R^2$  of 0.3%. Also as expected, and consistent with much prior research (e.g., Cawley, Heckman, & Vytacil, 2001), a 1-standard-deviation increase in cognitive ability is associated with 1.15 additional years of schooling ( $p < .001$ ,  $R^2 = 13\%$ ) in our dataset. Finally, in regression (8) with the human capital index as the outcome, the coefficient on genotype is reduced by including cognitive ability as a control, indicating that

cognitive ability is a statistical mediator. Applying the Sobel test for mediation (MacKinnon et al., 2002), we can reject the null hypothesis of no mediation ( $z = 3.37$ ,  $p = .0008$ ), and we estimate that cognitive ability mediates 51% of the relationship between the human capital index and the SNP.

The best test of whether a finding is a true positive is whether it replicates in multiple new, completely independent samples. Three additional research groups agreed to check in their data whether the association replicates: the Framingham Heart Study, the Wisconsin Longitudinal Study, and a sample of healthy control subjects for the Swedish Large Schizophrenia Association Study. Since we could not construct our human capital index in these samples, we studied only educational attainment, the most important component of the human capital index.

The Framingham Heart Study is a cardiovascular disease study that began in 1948 with a random sample of 5,209 participants from Framingham, Massachusetts. A sample with dense SNP data is available for 7,357 individuals, a mix of original participants and their relatives. Educational attainment is measured via nine categories, which we converted to estimated years of schooling. In this sample, educational attainment is *not* associated with SSADH rs2267539. In regression (8) with standardized years of education as the outcome, the coefficient on the number of A alleles is .06 with a standard error of .06 ( $p = .30$ ). While the SSADH SNP with the second-strongest association in the AGES-RS sample was not available in the Framingham Heart Study, we examined the SNP in the gene with the third-strongest association. The regression coefficient is .02 with a standard error of .06 ( $p = .70$ ). (In both cases, the standard errors are adjusted to correct for the presence of relatives.)

The Wisconsin Longitudinal Study is a random sample of 10,317 Wisconsin residents who graduated from high school in 1957, as well as 5,219 siblings who were enrolled later. We obtained genotypes for the three most significant SNPs in AGES-RS from a subsample of 3,408 individuals. Educational attainment is measured as years of schooling. Here it is not associated with any of our three most statistically significant SNPs from AGES-RS, and in fact the point estimates have the wrong sign in all three cases:  $\beta = -.02$  (s.e. = .07) for the most-strongly associated SNP;  $\beta = -.00$  (s.e. = .09) for the second SNP; and  $\beta = -.05$  (s.e. = .06) for the third SNP.

Our third non-Icelandic replication sample included 1,235 individuals from the healthy control group for the Swedish Large Schizophrenia Study. These are individuals who were identified from national population registers to match the schizophrenia group (which we do not analyze) along the characteristics of age, gender, and county of residence. Educational attainment is measured on a 1–6 scale, ranging from less than nine years to postgraduate education, which we convert to a standardized variable for the purposes of the regression analysis. It is not associated with either our most statistically significant SNP from AGES-RS ( $\beta = .06$ , s.e. = .07,  $p = .45$ ), nor with our second-most statistically significant SNP from AGES-RS ( $\beta = -.00$ , s.e. = .08,  $p = .98$ ).

What explains our puzzling pattern of results—the finding of an association that replicates with a sample similar to the original sample, passes various plausibility and robustness tests, and then fails to replicate in three other samples? We can think of four leading possibilities. First, the association in the AGES-RS data may be spurious due to confounding factors. For example, we attempted to deal with population stratification by controlling for the first two principal components of the whole-genome data, in addition to region dummies, an urban dummy, and region  $\times$  urban dummies. Even within an ethnically homogeneous population such as Iceland, however, there may be ethnic stratification on a finer scale than would be picked up by these controls. As a purely speculative example (meant just to illustrate the many possibilities), descendants of former nobility/leadership lineages could happen to have more A alleles and also be more educated. Second, the association may be a true positive, but local to the Icelandic environment. This could occur if, say, cognitive skills that are taught in schools outside of Iceland are instead self-taught within Iceland only by individuals with more A alleles. Third, the association may be a true positive, but local to the Icelandic genome, if the gene in question primarily has effects via its interaction with other genes and those genes differ between Icelanders and other populations. Fourth, our multiple hypothesis tests could have generated a false positive. Only due to chance did we happen to replicate the finding in a smaller sample from the AGES-RS data.

Patterns of results like ours are difficult to interpret. There are reasons to emphasize our replication within AGES-RS—which had exactly the same variable definitions and held constant the environmental and genotypic background—and discount our subsequent replication attempts: the ethnic makeup in the Framingham Heart Study and Wisconsin Longitudinal Study differ

substantially from the ethnic makeup in AGES-RS, and the Swedish sample, though ethnically more similar to the Icelandic one, is the smallest that we had. Yet there are also reasons to discount our plausibility and robustness checks. We chose our set of candidate genes *because* we thought they were most likely be involved in decision-making, so it is not surprising that the association we happened to find “makes sense” biologically. Cognitive ability is correlated with educational attainment, so it is not surprising that a SNP that happened to correlate within AGES-RS with educational attainment also correlates in that sample with cognitive function. Similarly, any confound that might explain the association between the human capital index and our most significant SNP on *SSADH* would equally well explain the association with other SNPs on *SSADH*, which are correlated with it.

Our failure to replicate a seemingly-robust association illustrates one of the major challenges for integrating genetics and social sciences. But our experience is not unique; indeed, it closely recapitulated a common storyline in medical genetics research.

### *Wisconsin Tale*

When we began our candidate gene study in AGES-RS, we believed—as did most medical-genetic researchers at the time—that the candidate gene approach was a reasonable approach. Our experience helped us to appreciate what had become, by the time our failure to replicate was complete, the new consensus view among the medical genetics community: The candidate gene approach is a path strewn with false positives.

These realizations made us skeptical of many published candidate gene associations. Yet social scientists, both authors and referees, seemed much less conscious of the fact that reported candidate gene associations are unlikely to be true. Consequently, we set out to systematically test existing candidate genes for general cognitive ability, also known as “intelligence,” or IQ, which is among the most highly studied psychological traits in molecular genetic work because it is among the most heritable of behavioral traits (estimates range from 0.50 to 0.80 for IQ measured in adulthood; Bouchard & McGue, 2003). There is a large literature of studies showing associations between many SNPs in various genes and general cognitive ability (see Payton, 2009 for a comprehensive review). As is typical of candidate gene studies in the social sciences, many of these results are based on small samples and had not seen any published replications.

As we report in Chabris et al. (forthcoming), we sought to replicate published associations between 12 genetic polymorphisms and general cognitive ability using three independent datasets—the previously-described Swedish Twin Registry, Wisconsin Longitudinal Study, and Framingham Heart Study, with a total sample size of 9,755 participants. Of 32 independent tests across all three datasets, only one was nominally significant at the  $p < .05$  level.<sup>13</sup> In the WLS data, where we tested all 12 genetic polymorphisms and had the most statistical power, we cannot reject the null hypothesis that the combined effect of those SNPs is zero—even though given our sample size of 5,571 individuals, we had 99% power to detect a combined effect of just  $R^2 = .52\%$ . Further power calculations suggested that, if the previously reported associations were not false positives, we should have expected between 11 and 15 replicated significant associations in our 32 tests, rather than the one that we found. Our analysis led us to conclude that most published SNP associations with general cognitive ability are false positives, most likely because the investigators in those studies inadvertently used samples that were much too small.<sup>14</sup>

### *Responding to the Inferential Challenges*

We believe there are several constructive responses to the inferential challenges posed by the small explanatory power of individual SNPs.

*Pooling data to increase power.* When it became widely recognized in the medical literature that candidate gene studies were generating a very high rate of false positives, and when dense SNP genotyping became sufficiently inexpensive, the standard research design became genome-wide association studies (GWAS). Obviously, relative to a candidate gene study, a GWAS magnifies the multiple-testing problem, but the stringent genome-wide significance threshold of  $p < 5 \times 10^{-8}$ , combined with implementing the GWAS in a large sample, has generated findings that have proven much more replicable.

Panel B of Table 3 shows the results of the same Bayesian calculation as Panel A, except for an association that is statistically significant at the  $p = 5 \times 10^{-8}$  level (rather than  $p = .05$ ).

---

<sup>13</sup> Intriguingly, the one nominally significant association was with *SSADH* rs2267539, the SNP implicated in our analysis of human capital in AGES-RS, but once again, the evidence is muddy due to multiple hypothesis testing.

<sup>14</sup> In other work using the Wisconsin Longitudinal Study on which we are collaborators, Freese et al. (2010) attempt to replicate associations reported in the literature between SNPs in the candidate gene *DRD2* and educational attainment, voting, partisanship, organization memberships, socializing, tobacco use, and alcohol use, and conclude that none of the associations replicate.

Because of the stringent significance threshold, statistical power is much lower at any given sample size  $N$ . Indeed, a true association with effect size  $R^2 = 0.1\%$  will probably *not* replicate at a genome-wide significance level for a sample smaller than 10,000 individuals, and there is only a 50% chance that an association known to be true will be detected in a sample of 30,000 individuals. Nonetheless, an association that reaches statistical significance at the genome-wide significance level in a sample of 10,000 or more individuals is almost certain to be a true positive.

Recognizing this, the medical literature has been moving in the direction of forming consortia of data providers. In such a consortium, a GWAS is conducted in each dataset, and the “discovery phase” is carried out as a meta-analysis of these GWAS results, a so-called *meta-GWAS*. In the “replication phase” that follows, the associations implicated in the discovery phase are investigated in independent samples.

On the one hand, the hurdle that genotype-outcome associations must hold in different samples that are typically drawn from populations with different ethnicities and environments—a requirement that is implicit in a meta-GWAS and explicit in the requirement that associations replicate in independent samples—means that GWAS researchers are unlikely to identify genetic associations that only exist in particular environments. The set of associations that are reported will tend to be ones that are among the strongest and the most universal—indeed, the fact that many true associations will not be discovered by a meta-GWAS perhaps helps to account for the “missing heritability” puzzle discussed in Section 3. On the other hand, the samples used in meta-GWAS have proven to be sufficiently large to detect SNP associations with modest effect sizes, and the findings that have emerged from these cooperative studies appear to be more likely to survive the challenges of replication.

Following the lead of the medical genetics community, we are organizing the Social Science Genetic Association Consortium (SSGAC), attempting to include all relevant major data providers that have dense SNP data and social science outcome measures. The SSGAC has had two meetings since its formation in February, 2011, under the auspices and guidance of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (Psaty et al. 2009), a successful medical genetic consortium. In forming the SSGAC, we are following in the footsteps of, and proceeding in close coordination with, the “Entrepreneurship

consortium” that was formed for the purpose of studying genetic associations with self-employment (van der Loos et al., 2010).

*Exploiting the cumulative effect of many SNPs.* Even in those cases where sample sizes are too small to discover robust associations, the data may still contain valuable information about the distribution of effect sizes of the SNPs on a dense SNP chip. Yang et al. (2010) developed a method, **Genomic REstricted Maximum Likelihood Estimation (GREML)**, for estimating the proportion of variance explained jointly by all the SNPs measured on a dense SNP chip. The key assumption is that among individuals who are unrelated—i.e., distantly related, since all humans are related to some extent—residual factors are uncorrelated with differences in the degree of genetic relatedness. Under that assumption, an estimate of heritability can be obtained by examining how the correlation in an outcome between pairs of individuals relates to the genetic distance between those individuals. Unlike in twin studies where relatedness is known, here the relatedness is directly estimated from the SNP data.

Unlike GWAS, for moderately heritable traits, GREML is well-powered for samples of only several thousand unrelated individuals because it aggregates the information contained in the genetic data. The GREML procedure estimates the fraction of variance of an outcome that could be predicted if a researcher had GWAS data and a look-up table that contained the *true* effect of all SNPs. Under the assumption that large individual-SNP effects are more likely for outcomes where the joint predictive power of all SNPs is larger, GREML can be used to assess which outcomes are the most promising to pursue for GWAS. Applying this method in a sample of 3,925 individuals, Yang et al. (2010) found that the measured SNPs could account for 45% of the variance in human height; Davies et al. (2011) apply the method to cognitive ability and obtain point estimates of 40% for crystallized intelligence ( $N = 3,254$ ) and 51% for fluid intelligence ( $N = 3,181$ ); and Chabris et al. (forthcoming) similarly estimate 47% for general cognitive ability.

As with GREML, the basic insight behind polygenic risk prediction (e.g., Purcell et al. 2009) is that even when it is not possible to robustly identify the individual SNPs associated with an outcome, it may still be possible to make statistically efficient use of the joint predictive power of a large number of SNPs. While GREML estimates the amount of predictive power theoretically attainable from the SNP data (but does not enable one to actually predict the outcome), a polygenic risk score is an attempt to use the SNP data in a given sample to actually

construct a predictive equation for an outcome in that sample. Naturally, in a finite sample, polygenic risk prediction will achieve much less predictive power than the theoretical bound estimated by GREML.

Unfortunately, the out-of-sample predictive power that can be obtained from considering the SNP data simultaneously is presently too small to be of practical use for most outcomes. For example, the International Schizophrenia Consortium (Purcell et al., 2009) reported an out-of-sample predictability of up to 3% from a predictive risk equation estimated in a total sample of 6,907 individuals. There is no reason to think that the predictability of economic traits will be greater when comparable sample sizes are available for constructing the prediction equation.

*Focusing on biologically proximal traits.* Regardless of analytic approach, a major question going forward is which outcomes are the most promising to study. In our view, in the short run, this decision will be dictated by which variables are consistently measured across a large enough number of datasets that the joint sample size will yield reliable results. For this reason, SSGAC's first outcome to study is educational attainment, which is widely measured not only in social science surveys, but also in most medical surveys as a key measure of socioeconomic background.

In the long run, however, we suspect that the most promising economic outcomes will be those that are most closely related to the underlying biology. Distal outcomes, such as educational attainment and self-employment, are likely influenced by an enormous number of genes, each with a tiny effect that will be difficult to detect even in a huge dataset. If these distal genetic effects work through different pathways in different local environments, then even true relationships will not robustly replicate across datasets. Proximal outcomes, such as aggressiveness and perhaps impulsivity, are likely to have larger and more direct genetic influences from fewer genes. Outcomes that are also measurable in animals have the additional advantage that the genes can be experimentally manipulated in animal models to directly study their causal effects. Unfortunately, as of now, none of these proximal outcomes are widely measured across many datasets that have dense SNP data. One function we envision for the SSGAC will be to coordinate the funding and collection of harmonized measures of proximal outcomes.



## 5. Conclusion: Genomics research in economics

We have discussed a number of ways in which the use of molecular genetic data could benefit economics. Genetic data will serve as a powerful lens to identify and study biological mechanisms that generate important, and potentially overlooked, sources of inter-individual differences (e.g., aggression, ambition, myopia, etc.). Genotypic data will also (eventually) be used as control variables that serve to increase power. Genetic data may also be of interest in and of itself: economists have used genotypic data to study the effect of intellectual property rights on innovation (Williams, 2010) and adverse selection in health insurance markets (Oster et al., 2010). Looking ahead a decade or two, the availability of inexpensive genotypic data is likely to help parents predict learning disabilities like dyslexia earlier in childhood, facilitating earlier interventions. Potential vulnerabilities to substance abuse, or other kinds of self-destructive behavior, may also one day be predictable from genetic data.

We also predict that methodological challenges—such as multiple testing—will generate many more false positives in the literature, especially in the short run. The press is likely to distort findings and exaggerate the degree to which specific genes “determine” outcomes. In most cases there is no “gene for [insert behavior here],” despite what the newspaper headlines say. Indeed, for most behaviors researchers are struggling to find a SNP with an  $R^2$  that is greater than 1/10 of 1%. Researchers in this field hold a special responsibility to try to accurately inform the media and the public about the limitations of the science.

The inevitable, inexpensive, broad-based availability of genotypic information will raise myriad social, ethical, and legal questions to which economic analysis will provide a valuable perspective. Many geneticists rightly worry that genetic research will prove to be socially harmful by generating discrimination against genetically-disadvantaged groups. Genetic information will generate a rich set of new policy problems (in addition to the benefits that we reviewed above). Governments will need to formulate new policies that maximize social welfare in a world where people with genetic advantages will wish to share them with potential employers and insurers, and people with genetic disadvantages will want to shroud them. In some cases, provision of genetic information can be beneficial (e.g., alerting couples who both possess disease-causing recessive mutations), while in other cases, it would be deeply problematic (e.g., sharing genetic data with health insurance companies, which effectively creates an unraveling of some of the social benefit of health insurance). Problems abound in any

analysis of optimal access to genetic information, even when the individual herself is the only one who is going to have access to the data. Under what conditions will the benefits to an individual from knowing her own genetic risk factors, such as the ability to prepare well in advance for a likely illness, outweigh the costs of increased anxiety and distress (cf. Oster, Dorsey, & Shoulson, 2011)? We predict that research on these different types of questions will soon occupy a much larger fraction of economists' energy as these issues quickly become of immediate practical relevance.

### References

- Apicella CL, Cesarini D, Johannesson M, Dawes CT, Lichtenstein P, Wallace B, Beauchamp J, Westberg L. forthcoming. No association between oxytocin receptor (OXTR) gene polymorphisms and experimentally elicited social preferences. *PLoS ONE*.
- Barnea A, Cronqvist H, Siegel S. 2010. Nature or Nurture: What Determines Investor Behavior? *Journal of Financial Economics*, 98(3): 583–604
- Barnett JH, Scoriels L, Munafò MR. 2008. Meta-analysis of the cognitive effects of the catechol-O-methyltransferase gene Val158/108Met polymorphism. *Biological Psychiatry*. 64: 137-144
- Beauchamp JP, Cesarini D, Johannesson M, der Loos M, Koellinger P, Groenen PJF, Fowler J, Rosenquist N, Thurik R, Christakis N. 2011a. Molecular genetics and economics. *Journal of Economic Perspectives*. 25(4): 1-27.
- Beauchamp JP, Cesarini D, Johannesson M. 2011b. The Psychometric Properties of Measures of Economic Preferences. Harvard University Manuscript
- Becker GS. 1993. Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy*. 101(3): 385-409
- Becker GS, Tomes N. 1976. Child Endowments and the Quantity and Quality of Children. *Journal of Political Economy*. 84(no. 4, part 2): S143-62
- Becker GS, Tomes N. 1986. Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*. 4 (no. 3, pt. 2): S1-S39
- Benjamin DJ, Chabris CF, Glaeser EL, Gudnason V, Harris T, Laibson DI, Launer L, Purcell S. 2007. Genoeconomics in *Biosocial surveys*, ed. M Weinstein, JW Vaupel, KW Wachter; pp. 304–335. Washington, DC: The National Academies Press
- Billings LK, Florez JC. 2010. The genetics of type 2 diabetes: what have we learned from GWAS? *Annals of the New York Academy of Science*. 1212: 59-77
- Björklund A, Jäntti M, Solon G. 2005. Influences of Nature and Nurture on Earnings Variation: A Report on a Study of Various Sibling Types in Sweden in *Unequal Chances: Family*

- Background and Economic Success*, ed. S Bowles, H Gintis, M Osborne Groves. Princeton, NJ: Princeton University Press.
- Bouchard TJ, McGue M. 2003. Genetic and Environmental Influences on Human Psychological Differences. *Journal of Neurobiology*. 54(1): 4-45
- Buzzi AYW, Frantseva MV, Velazquez JLP, Cortez MA, Liu CC, Shen LQ, Gibson KM, Snead OC. 2006. 3rd. Succinic semialdehyde dehydrogenase deficiency: GABAB receptor-mediated function. *Brain Research*. 1090(1): 15-22
- Carpenter J, Garcia J, Loji Lum J. 2011. Dopamine Receptor Genes Predict Risk Preferences, Time Preferences, and Related Economic Outcomes. *Journal of Risk and Uncertainty*. 42(3):233-261
- Cawley J, Heckman J, Vytlačil E. 2001. Three observations on wages and measured cognitive ability. *Labour Economics*. 8: 419-442
- Cecil JE, Tavendale R, Watt P, Hetherington MM, Palmer CAN. 2008. An Obesity-Associated FTO Gene Variant and Increased Energy Intake in Children. *New England Journal of Medicine*. 359: 2558-2566
- Cesarini D, Dawes CT, Fowler J, Johannesson M, Lichtenstein P, Wallace B. 2008. Heritability of Cooperative Behavior in the Trust Game. *Proceedings of the National Academy of Sciences*. 105: 3271-3276
- Cesarini D, Dawes CT, Johannesson M, Lichtenstein P, Wallace B. 2009. Genetic Variation in Preferences for Giving and Risk-Taking. *Quarterly Journal of Economics*. 124(2): 809-842
- Cesarini D, Johannesson M, Lichtenstein P, Sandewall O, Wallace B. 2010. Genetic Variation in Financial Decision Making. *Journal of Finance*. 65(5): 1725-1754
- Cesarini D, Johannesson M, Magnusson PKE, Wallace B. forthcoming. The Behavioral Genetics of Behavioral Anomalies. *Management Science*
- Chabris CF, Hebert BM, Benjamin DJ, Beauchamp J, Cesarini D, Johannesson M, Magnusson PKE, Lichtenstein P, Atwood CS, Freese J, Hauser TS, Hauser RM, Christakis N, Laibson D. forthcoming. Most Published Genetic Associations with General Cognitive Ability Are False Positives. *Psychological Science*
- Chen L, Davey Smith G, Harbord R, Lewis S. 2008. Alcohol intake and blood pressure: a systematic review implementing Mendelian Randomization approach. *PLoS Medicine*. 5(3): 461-471
- Conley D. 2009. The Promise and Challenges of Incorporating Genetic Data into Longitudinal Social Science Surveys and Research. *Biodemography and Social Biology*. 55: 238-251
- Crisan LG, Pana S, Vulturar R, Heilman RM, Szekely R, Druga B, Dragos N, Miu AC. 2009. Genetic contributions of the serotonin transporter to social learning of fear and economic decision making. *Social Cognitive and Affective Neuroscience*. 4: 399-408

- Davey Smith G, Ebrahim S. 2003. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*. 32: 1-22
- Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, Ke X, Le Hellard S, Christoforou A, Luciano M, McGhee K, Lopez L, Gow AJ, Corley J, Redmond P, Fox HC, Haggarty P, Whalley LJ, McNeill G, Goddard ME, Espeseth T, Lundervold AJ, Reinvang I, Pickles A, Steen VM, Ollier W, Porteous DJ, Horan M, Starr JM, Pendleton N, Visscher PM, Deary IJ. 2011. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*. 16: 996-1005
- de Moor MHM et al. 2012. Meta-analysis of genome-wide association studies for personality. *Molecular Psychiatry*. 17: 337-349
- De Neve J-E, Fowler JH. 2010. The MAOA Gene Predicts Credit Card Debt. University College London Manuscript
- De Neve J-E, Fowler JH, Frey BS, Christakis NA. 2011. Genes, Economics, and Happiness. University College London Manuscript
- De Rango F, Leone O, Dato S, Novelletto A, et al. 2008. Cognitive functioning and survival in the elderly: the SSADH C538T polymorphism. *Annals of Human Genetics*. 72(5): 630-635
- Ding W, Lehrer S, Rosenquist N, Audrain-McGovern J. 2009. The Impact of Poor Health on Academic Performance: New Evidence Using Genetic Markers. *Journal of Health Economics*. 28(3): 578-597
- Dreber A, Apicella CL, Eisenberg DTA, Garcia JR, Zamore R, Lum JK, Campbell BC. 2009. The 7R Polymorphism in the Dopamine Receptor D4 Gene (DRD4) is associated with financial risk-taking in men. *Evolution and Human Behavior*. 30(2): 85-92
- Dreber A, Rand DG, Wernerfelt N, Garcia JR, Vilar MG, Lum JK, Zeckhauser RJ. forthcoming. Dopamine and Risk Choices in Different Domains: Findings Among Serious Tournament Bridge Players. *Journal of Risk and Uncertainty*
- Duncan LE, Keller MC. 2011. A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. *American Journal of Psychiatry*. 168(10): 1041-1049
- Ebstein RP, Israel S, Chew SH, Zhong S, Knafo A. 2010. Genetics of human social behavior. *Neuron*. 65(6): 831-844
- Eisenberg DT, Mackillop J, Modi M, Beauchemin J, Dang D, Lisman SA, Lum JK, Wilson DS. 2007. Examining impulsivity as an endophenotype using a behavioral approach: a DRD2 TaqI A and DRD4 48-bp VNTR association study. *Behavioral and Brain Functions*. 3: 2
- Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*. UK: Benjamin Cummings

- Fletcher J, Lehrer S. 2009. The Effects of Adolescent Health on Educational Outcomes: Causal Evidence Using Genetic Lotteries between Siblings. *Forum for Health Economics & Policy*, 12(2), Health and Education, Article 8
- Frayling TM, Timpson NJ, Weedon MN, et al. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 316: 889-94
- Freese J, Branigan AR, Atwood CS, Hauser TS, Benjamin DJ, Chabris CF, Laibson DI, Hauser RM. 2010. *Taq1a* and College Attendance, Partisanship, Voting, and Other Outcomes: Replication Attempts Using The Wisconsin Longitudinal Study. Northwestern University Manuscript
- Frydman C, Camerer C, Bossaerts P, Rangel A. 2011. MAOA-L carriers are better at making optimal financial decisions under risk. *Proceedings of the Royal Society*. 278(1714): 2053-2059
- Goldberger A. 1979. Heritability. *Economica*. 46: 327-347
- Hammock EAD, Young LJ. 2002. Variation in the vasopressin V1a receptor promoter and expression: Implications for inter- and intra- specific variation in social behavior. *European Journal of Neuroscience*. 16: 399-402
- Hammock EA, Lim NM, Nair HP, Young LJ. 2005. Association of vasopressin 1a receptor levels with a regulatory microsatellite and behavior. *Genes, Brain and Behavior*. 4(5): 289–301
- Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, Sigurdsson G, Thorgeirsson G, Aspelund T, Garcia ME, Cotch MF, Hoffman HJ, Gudnason V. 2007. Age, Gene/Environment Susceptibility-Reykjavik Study: Multidisciplinary Applied Phenomics. *American Journal of Epidemiology*. 165(9): 1076–1087
- Hewitt JK. 2011. Editorial Policy on Candidate Gene Association and Candidate Gene-by-Environment Interaction Studies of Complex Traits. *Behavior Genetics*. Published online September 18
- Hirschhorn JN. 2009. Genomewide Association Studies—Illuminating Biologic Pathways. *New England Journal of Medicine*. 360(17): 1699–1701
- Israel S, Lerer E, Shalev I, Uzefovsky F, Riebold M, Laiba E, Bachner-Melman R, Maril A, Bornstein G, Knafo A, Ebstein R. 2009. The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the dictator game and the social value orientations task. *PLoS One*. 4: e5535
- Jang KL, Livesley WJ, Vernon PA. 1996. Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study. *Journal of Personality*. 64(3): 577-59.
- Jencks CS. 1980. Heredity, environment, and public policy reconsidered. *American Sociological Review*. 45(5): 723-736
- Knafo A, Israel S, Darvasi A, Bachner-Melman R, Uzefovsky F, Cohen L, Feldman E, Lerer E, Laiba E, Raz Y, Nemanov L, Gritsenko I, Dina C, Agam G, Dean B, Bornstein G, Ebstein RP. 2008. Individual differences in allocation of funds in the Dictator Game and postmortem hippocampal mRNA levels are correlated with length of the arginine

- vasopressin 1a receptor (AVPR1a) RS3 promoter-region repeat. *Genes, Brain and Behavior*. 7: 266-275
- Knerr I, Gibson KM, Jakobs C, Pearl PL. 2008. Neuropsychiatric morbidity in adolescent and adult succinic semialdehyde dehydrogenase deficiency patients. *CNS Spectrum*. 13(7): 598-605
- Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. 2005. Oxytocin increases trust in Humans. *Nature*. 435: 673–676
- Kuhnen, CM., and JY Chiao. 2009. Genetic Determinants of Financial Risk Taking. *PLoS ONE*. 4(2), e4362
- Kuhnen C, Samanez-Larkin GR, Knutson B. 2011. Serotonin and Risk Taking: How Do Genes Change Financial Choices? Kellogg School of Management Manuscript
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. 2001. Initial Sequencing and Analysis of the Human Genome. *Nature*. 409(6822): 860-921
- Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. *Science*. 265: 2037-2948
- Lango Allen H, et al. 2010. Hundreds of Variants Clustered in Genomic Loci and Biological Pathways Affect Human Height. *Nature*. 467(7317): 832–838
- Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Smith GD. 2008. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*. 27: 1133-1163
- Lee JJ. 2010. Review of intelligence and how to get it: Why schools and cultures count, R.E. Nisbett, Norton, New York, NY (2009). *Personality and Individual Differences*. 48(2): 247-255
- Leone O, et al. 2006. A human derived SSADH coding variant is replacing the ancestral allele shared with primates. *Annals of Human Biology*. 33: 593-603
- Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, Pedersen NL. 2002. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *Journal of Internal Medicine*. 252: 184–205
- Lizzeri A, and Siniscalchi M. 2008. Parental Guidance and Supervised Learning. *Quarterly Journal of Economics*. 123(3): 1161-1195
- MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. 2002. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*. 7: 83-104
- Manski C. 2011. Genes, Eyeglasses, and Social Policy. *Journal of Economic Perspectives*. 25(4): 83-94
- Martin NW, Medland SE, Verweij KJH, Lee SH, Nyholt DR, Madden PA, Heath AC, Montgomery GW, Wright MJ, Martin NG. 2011. Educational Attainment: A Genome Wide Association Study in 9538 Australians. *PLoS ONE*. 6(6): e20128

- Mazumder B. 2005. The apple falls even closer to the tree than we thought: New and revised estimates of the intergenerational transmission of earnings in *Unequal Chances: Family Background and Economic Success*, ed. S Bowles, H Gintis, M Osborne Groves. Princeton, NJ: Princeton University Press
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP A, et al. 2008. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Review Genetics*. 9: 356–369
- McDermott R, Tingley D, Cowden J, Frazzetto G, Johnson D. 2009. Monoamine Oxidase A gene (MAOA) predicts behavioral aggression following provocation. *Proceedings of the National Academy of Sciences*. 106(7): 2118-2123
- Nicolaou N, Shane S, Adi G, Mangino M, Harris J. 2011. A Polymorphism Associated with Entrepreneurship: Evidence from Dopamine Receptor Candidate Genes. *Small Business Economics*. 36(2): 151-55
- Norton EC, Han E. 2008. Genetic information, obesity, and labor market outcomes. *Health Economics*. 17(9): 1089-1104
- Obeidat M, et al. 2011. A Comprehensive Evaluation of Potential Lung Function Associated Genes in the SpiroMeta General Population Sample. *PLoS ONE*. 6(5): e19382
- Oster E, Shoulson I, Quaid KA, Dorsey ER on behalf of the Huntington Study Group PHAROS Investigators. 2010. Genetic adverse selection: evidence from long-term care insurance and Huntington disease. *Journal of Public Economics*. 94: 1041-1050
- Oster E, Shoulson I, Dorsey ER. 2011. Optimal expectations and limited medical testing: Evidence from Huntington disease. November, University of Chicago Manuscript
- Payton A. 2009. The impact of genetic research on our understanding of normal cognitive ageing: 1995 to 2009. *Neuropsychology Review*. 19(4): 451-477
- Pence KM. 2006. The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving. *Contributions to Economic Analysis & Policy*. 5(1): Article 20
- Pinker S. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking
- Plomin R, Owen MJ, McGuffin P. 1994. The genetic basis of complex human behaviors. *Science*. 264: 1733–173
- Plomin R, Turic DM, Hill L, Turic DE, Stephens M, Williams J, Owen MJ, O'Donovan MC. 2004. A functional polymorphism in the succinate-semialdehyde dehydrogenase (aldehyde dehydrogenase 5 family, member A1) gene is associated with cognitive ability. *Molecular Psychiatry*. 9: 582-586
- Plomin R, DeFries JC, McClearn GE, McGuffin P. 2008. *Behavioral Genetics*. New York: Worth Publishers

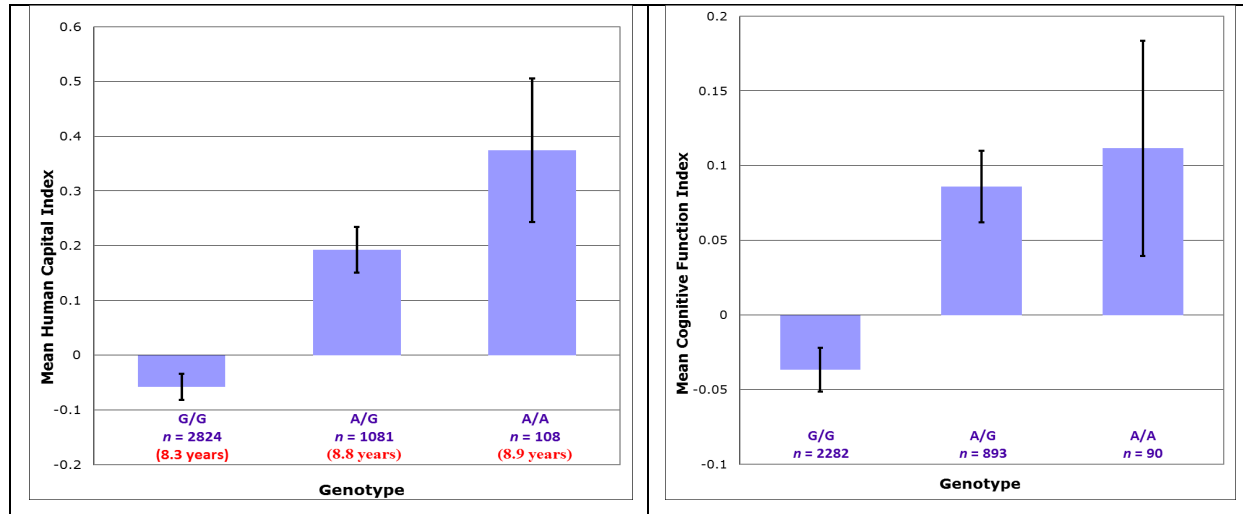
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. 2006. Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. *Nature Genetics*. 38(8): 904-909
- Price AL, Helgason A, Palsson S, Stefansson H, St. Clair D, Andreassen OA, Reich D, Kong A, Stefansson K. 2009. The Impact of Divergence Time on the Nature of Population Structure: An Example from Iceland. *PLoS Genetics*. 5(6):1-10
- Psaty B, et al. 2009. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. Design of prospective meta-analysis of genome-wide association studies from 5 cohorts. *Circulation: Cardiovascular Genetics*. 2:73-80
- Purcell S, et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(6), 748-752
- Roe BE, et al. 2010. Financial and Psychological Risk Attitudes Associated with Two Single Nucleotide Polymorphisms in the Nicotine Receptor (*CHRNA4*) Gene. *PLoS ONE*. 4(8): e6704
- Rosenquist JN, O'Malley AJ, Lehrer SF, Zaslavsky A, Smoller JW, Christakis NA. 2012. Genotype-Phenotype Association of FTO with Body Mass Index is Modified by Birth Era. Harvard University Manuscript
- Rowe DC, Vesterdal WJ, Rodgers JL. 1999. Herrnstein's syllogism: genetic and shared environmental influences on IQ, education and income. *Intelligence*, 26: 405-423
- Sacerdote, B. 2010. Nature and Nurture Effects on Children's Outcomes: What Have We Learned from Studies of Twins and Adoptees? In *Handbook of Social Economics*, vol 1A, ed. Jess Benhabib, Matthew Jackson, and Alberto Bisin, 1–30. Amsterdam: North Holland
- Schumacher J, Hoffman P, Schmääl C, Schulte-Körne G, Nöthen M. 2007. Genetics of dyslexia: The evolving landscape. *Journal of Medical Genetics*. 44: 289-297
- Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, DeLange M, Harris JR, Hjelmborg JV, Luciano M, Martin NG, Mortensen J, Nisticò L, Pedersen NL, Skytthe A, Spector TD, Stazi MA, Willemsen G, Kaprio J. 2003. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Research and Human Genetics*. 6(5): 399-408
- Solon GR. 1992. Intergenerational Income Mobility in the United States. *American Economic Review*. 82(3): 393-408
- St. George-Hyslop PH. 2000. Molecular genetics of Alzheimer's disease. *Biological Psychiatry*. 47: 183–199
- Stenberg A. 2011. Nature or Nurture? A Note on the Misinterpreted Twin Decomposition. Work. Pap. 4/2011, Swedish Institute for Social Research (SOFI)



- Strittmatter WJ, Saunders AM, Schmechel D, et al. 1993. Apolipoprotein E: High avidity binding to  $\beta$ -amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences*. 90: 1977–1981
- Taubman P. 1976. The Determinants of Earnings: Genetics, Family, and Other Environments: A Study of White Male Twins. *American Economic Review*. 66: 858-870
- Tung YC, Yeo GS. 2011. From GWAS to biology: lessons from FTO. *Annals of the New York Academy of Science*. 1220: 162-171
- Turkheimer E. 2000. Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*. 9: 160–164
- van der Loos MJHM, Koellinger PD, Groenen PJF, Thurik AR 2010. Genome-Wide Association Studies and the Genetics of Entrepreneurship. *European Journal of Epidemiology*. 25(1): 1-3
- van der Loos MJHM, Koellinger PD, Groenen PJF, Rietveld CA, Rivadeneira F, van Rooij FJA, Uitterlinden AG, Hofman A, Thurik AR. 2011. Candidate gene studies and the quest for the entrepreneurial gene. *Small Business Economics*. 37: 269-275
- Venter CJ, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The Sequence of the Human Genome. *Science*. 291: 1304-1351
- von Hinke Kessler Scholder S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. 2010. Genetic Markers as Instrumental Variables: An Application to Child Fat Mass and Academic Achievement. Work. Pap. 10/229, University of Bristol
- Visscher PM, Hill WG, Wray NR. 2008. Heritability in the genomics era — Concepts and misconceptions. *Nature Reviews Genetic*. 9: 255–266
- Wallace B, Cesarini D, Lichtenstein P, Johannesson M. 2007. Heritability of Ultimatum Game Responder Behavior. *Proceedings of the National Academy of Sciences*. 104: 15631-15634
- Williams H. 2010. Intellectual Property Rights and Innovation: Evidence from the Human Genome. Work. Pap. #16213, NBER
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. 2010. Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nature Genetics*. 42(7): 565-569
- Zhong S, Chew SH, Set E, Zhang J, Xue H, Sham PC, Ebstein RP, Israel S. 2009a. The Heritability of Attitude Toward Economic Risk. *Twin Research and Human Genetics*. 12(1): 103-107
- Zhong S, Israel S, Xue H, Sham PC, Ebstein RP, Chew SH. 2009b. A neurochemical approach to valuation sensitivity over gains and losses. *Proceedings of the Royal Society B*. 276: 4181–418.
- Zhong S, Israel S, Xue H, Ebstein RP, Chew SH. 2009c. Monoamine oxidase A gene (MAOA) associated with attitude towards longshot risks. *PLoS ONE*. 4(12): e851.

- Zhong S, Israel S, Xue H, Ebstein RP, Chew SH. 2010. Dopamine D4 receptor gene associated with fairness preference in ultimatum game. *PLoS ONE*. 5(11): e13765
- Zimmerman D. 1992. Regression toward Mediocrity in Economic Stature. *American Economic Review*. 82: 409-429
- Zyphur M, Narayanan J, Arvey R, Alexander G. 2009. The Genetics of Economic Risk Preferences. *Journal of Behavioral Decision Making*. 22(4): 367-377

**Figure 1. Mean of human capital index by genotype**



Notes: Data are from AGES-RS. Human capital index is a composite variable made up of educational attainment and number of languages learned. In the human capital index panel, mean years of educational attainment by genotype are shown in parentheses below the sample size. Cognitive function index is a composite variable made up of digit symbol substitution (WAIS), digit span (forward and backward), spatial working memory, long-term memory (CVLT recall and recognition). In the cognitive function sample, survey respondents who scored  $\leq 23$  on the Mini Mental State Exam are dropped. Both the human capital index and the cognitive function index are standardized to have zero mean and unit variance. Genotype is for *SSADH* rs2267539. Error bars show  $\pm 1$  standard error.

**Table 1: Sibling Correlations for Log-Income Averaged over Multiple Years**

	Men			Women		
	<u>MZ</u>	<u>DZ</u>	<u>p-value</u>	<u>MZ</u>	<u>DZ</u>	<u>p-value</u>
1 Year	<i>0.406</i> (0.344- 0.474)	<i>0.164</i> (0.123- 0.211)	<0.001	<i>0.266</i> (0.206- 0.327)	<i>0.143</i> (0.095- 0.190)	0.002
3 Year	<i>0.513</i> (0.412- 0.633)	<i>0.193</i> (0.150- 0.243)	<0.001	<i>0.293</i> (0.237- 0.352)	<i>0.137</i> (0.099- 0.184)	<0.001
5 Year	<i>0.512</i> (0.447- 0.574)	<i>0.201</i> (0.161- 0.251)	<0.001	<i>0.297</i> (0.239- 0.361)	<i>0.198</i> (0.145- 0.253)	<0.001
10 Year	<i>0.556</i> (0.486- 0.618)	<i>0.241</i> (0.199- 0.285)	<0.001	<i>0.353</i> (0.293- 0.419)	<i>0.226</i> (0.180- 0.272)	<0.001
20 Year	<i>0.626</i> (0.574- 0.676)	<i>0.270</i> (0.223- 0.317)	<0.001	<i>0.481</i> (0.431- 0.528)	<i>0.221</i> (0.170- 0.282)	<0.001

Notes: Data are from the SALT sample of the Swedish Twin Registry. This table reports the log-income correlations for MZ and DZ twin pairs, separately by sex, with log-income averaged over 1, 3, 5, 10, and 20 years. Income is defined as the sum of income earned from wage labor, income from own business, pension income and unemployment compensation. The sample is restricted to those individuals for whom there is no income data at ages 31 through 50 and the average income exceeds 1000 SEK (approximately \$150). Confidence intervals are in parentheses below the point estimates. Confidence intervals and *p*-values are bootstrapped.

**Table 2. Sibling Correlations for Wealth**

	Men			Women			Pooled		
	<u>MZ</u>	<u>DZ</u>	<u>p-value</u>	<u>MZ</u>	<u>DZ</u>	<u>p-value</u>	<u>MZ</u>	<u>DZ</u>	<u>p-value</u>
Hyperbolic Sine	<i>0.088</i> (-.015-.249)	<i>0.162</i> (-.007-.375)	0.692	<i>0.446</i> (0.181-.573)	<i>0.301</i> (.079-.280)	0.019	<i>0.282</i> (.144-.434)	<i>0.109</i> (.013-.256)	0.054
Standard Normal	<i>0.411</i> (.277-.533)	<i>0.378</i> (.220-.524)	0.381	<i>0.380</i> (.309-.569)	<i>0.072</i> (.152-.460)	0.096	<i>0.432</i> (.331-.529)	<i>0.336</i> (.219-.440)	0.084

Notes: Data are from the first wave of the SALTY sample of the Swedish Twin Registry. Net wealth is defined as the difference between the total self-reported value of assets and total self-reported debt. The asset classes considered are: property (including summer house), stocks, bonds, transportation vehicles and other. Respondents are asked to prorate in cases of joint ownership of an asset or joint debt. The exact question wording is in Online Appendix I. Net wealth is transformed as described in the text. The results do not change appreciably if we remove outliers by restricting the sample to individuals with an absolute net wealth lower than 10,000,000 SEK (approximately \$1,500,000). Confidence intervals are in parentheses below the point estimates. Confidence intervals and *p*-values are bootstrapped.

**Table 3. Posterior probability of a true association of  $R^2 = 0.1\%$  as a function of prior probability and sample size ...**

**Panel A: ... for an association that is statistically significant at  $p = .05$ :**

		Sample size				
		$N = 100$	$N = 1,000$	$N = 5,000$	$N = 10,000$	$N = 30,000$
		(power = .06)	(power = .17)	(power = .61)	(power = .89)	(power = .99)
Prior	.01%	.01%	.03%	.12%	.18%	.20%
probability	1%	1%	3%	11%	15%	17%
of true	10%	12%	27%	58%	66%	69%
association						

**Panel B: ... for an association that is statistically significant at  $p = 5 \times 10^{-8}$ :**

		Sample size				
		$N = 100$	$N = 1,000$	$N = 5,000$	$N = 10,000$	$N = 30,000$
		(power = .00)	(power = .00)	(power = .00)	(power = .01)	(power = .51)
Prior	.01%	.03%	3%	57%	96%	100%
probability	1%	3%	47%	99%	100%	100%
of true	10%	25%	91%	100%	100%	100%
association						

Notes: See text for the assumptions underlying these calculations. Power is calculated using Purcell, Cherny, and Sham's (2003) online tool: <http://pngu.mgh.harvard.edu/~purcell/gpc/qtlassoc.html>.

Posterior probabilities are then calculated by Bayes' Rule:

$$\text{Pr}(\text{true}|\text{significant}) = (\text{power} \times \text{prior}) / ((\text{power} \times \text{prior}) + (.05 \times (1 - \text{prior}))).$$

**Table 4. OLS regression of human capital index on genotype**

	(1)	(2)	(3)	(4)
Genotype (#A alleles)	0.218 (0.054)	0.178 (0.060)	0.185 (0.059)	0.187 (0.059)
Birth year	0.055 (0.005)	0.056 (0.006)	0.046 (0.006)	0.046 (0.006)
Female	-0.692 (0.056)	-0.681 (0.062)	-0.682 (0.061)	-0.684 (0.062)
Urban			1.054 (0.242)	1.323 (0.578)
GWAS Principal Comp.?	NO	YES	YES	YES
Region FE?	NO	NO	YES	YES
Region × Urban FE?	NO	NO	NO	YES
$R^2$	0.068	0.069	0.100	0.102
$N$	4,016	3,198	3,198	3,198

Notes: Data are from AGES-RS. Human capital index is a composite variable made up of educational attainment and number of languages learned. It is standardized to have zero mean and unit variance. Genotype is for *SSADH* rs2267539. Coefficients for the constant term and control variables are suppressed. Standard errors are in parentheses. Urban is a dummy variable for whether the respondent grew up in an urban area. GWAS Principal Components refers to the first 2 principal components of the dense SNP data. Region fixed effects are dummies for the nine regions of Iceland (see Price et al., 2009). The first column includes the 2,349 AGES-RS respondents whom we had genotyped with our SNP custom microarray, plus the non-overlapping subset of 3,198 AGES-RS respondents for whom dense SNP data were available. Because the other columns control for principal components of the dense SNP data, they only include the 3,198 respondents for whom dense SNP data were available.

# Glossary

***Candidate gene*** – A genetic polymorphism that is hypothesized to have a causal effect on some trait (or disease). The hypothesis is either based on what is believed about the biological function of the gene where the genetic polymorphism is located, or it is based on previously-reported associations between that genetic polymorphism and a related outcome.

***Gene*** – A sequence of nucleotides in DNA that provides instructions for building a particular protein or proteins.

***Genetic polymorphism*** – A segment of DNA that differs between individuals.

***Genome-wide association study (GWAS)*** – A study in which hundreds of thousands of genetic polymorphisms are individually tested for association with some outcome, without any prior hypotheses.

***Genome-wide significance*** –  $5 \times 10^{-8}$ ; the conventional level at which an association is considered to be statistically significant in a genome-wide association study.

***Genotype*** – For a given SNP, an individual's number of minor alleles.

***Major allele*** – The nucleotide of a SNP that is more common in the population. (For non-SNP genetic polymorphisms with two alleles, *major allele* is again the term for the more common variant in the population.)

***Meta-GWAS*** – A meta-analysis of results from multiple genome-wide association studies (GWAS).



***Minor allele*** – The nucleotide of a SNP that is less common in the population. (For non-SNP genetic polymorphisms with two alleles, *minor allele* is again the term for the less common variant in the population.)

***Pleiotropy*** – Multiple effects of a single gene.

***Single-nucleotide polymorphism (SNP)*** – A single nucleotide location in the DNA that varies across individuals.

# Online Appendix I

## Sample Description

**Sample for income analyses:** In our analysis of permanent income, we focus on twins in the so-called SALT, or Screening Across the Lifespan, cohort. SALT was an attempt, initiated in the 1990s, to screen all twins born between 1926 and 1958 by telephone. It attained a response rate of 71%. For further details on the sample, see Lichtenstein et al. (2002). We use the same algorithm as the Swedish Twin Registry to determine zygosity. Lichtenstein et al. (2002) discuss the method and its high level of reliability. Our sample also includes some twins who declined to take part in SALT but have answered other questionnaires administered by the STR and for which there is sufficient information to establish zygosity with high confidence.

When computing multi-year averages, we exclude years in which the income is less than the value of 1,000 dollars in 1985. We use the consumer price index of Statistics Sweden to establish the relevant threshold in other years. Table A1 provides descriptive statistics for the income variables we use, disaggregated by gender and birth cohort.

**Sample for the wealth analyses:** Our analyses of wealth uses data from a more recent survey called SALTY, the first major survey of twins which features entire sections specifically devoted to economic preferences and behaviors. Beginning in early 2009, the survey was sent out to 24,914 Swedish twins born between 1943 and 1958 for whom at least one member had participated in SALT. The data collection was completed in the summer of 2010. The survey generated a total of 11,418 usable responses, with a response rate of almost 50%. The wealth questions we studied here, however, were only administered to approximately 40% of the survey respondents. Cesarini et al. (2012) analyze non-response in the SALTY sample and find that respondents are more likely to be female, have higher education levels, and higher income than the non-responders. The differences are quite small, however.

Some summary statistics on the untransformed wealth measure are given in Table A2. We note that the standard deviations and means of the untransformed variable are extremely

sensitive to a small number of outlier observations, whose influence is greatly diminished when we work with the transformed wealth variables, as in the analysis in the text.

## Variable Description

**Net Wealth.** Net wealth is defined as the difference between gross wealth and debt.

Gross wealth is measured by summing the responses to the following questions.

Below, please state the value of your assets in each of the categories. By value we mean the value at which you could sell the assets tomorrow, if the need were to arise. State the value in SEK. If the assets are jointly owned with someone else, for example a spouse, only state your share of the assets.

- \_ Property, including summer cottage, forest and farmland
- \_ Stocks
- \_ Bonds
- \_ Bank
- \_ Boat, car and other vehicles.
- \_ Other assets, for example jewelry, antiques and art.

Debt is measured using the response to the question:

How large are your loans? Enter “0” if you do not have any loans.

**Income.** The income measure used in this is paper (sammanräknad förvärvsinkomst) is defined as the sum of income earned from wage labor, income from own business, pension income and unemployment compensation. Capital income is not included and the variables are not censored. Since administrative records only contain information on legally earned, taxed income, annual income is only an imperfect proxy for actual income earned.

## References

- Cesarini D, Johannesson M, Magnusson PKE, Wallace B. forthcoming. The Behavioral Genetics of Behavioral Anomalies. *Management Science*
- Lichtenstein P, et al. 2002. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies . *Journal of Internal Medicine* 252(3): 184–205

**Table A1: Summary Statistics for Income Averaged over Ages 31 to 50**

		Men		Women	
		<u>MZ</u>	<u>DZ</u>	<u>MZ</u>	<u>DZ</u>
1938-1943	Mean	105,865	102,822	64,771	62,219
	S.D.	56,362	47,983	27,313	27,953
	#pairs	517	795	516	731
1944-1949	Mean	153,028	150,767	100,155	95,726
	S.D.	63,678	74,664	39,174	40,166
	#pairs	645	994	730	1081
1950-1955	Mean	208,977	206,094	139,775	138,253
	S.D.	101,907	152,206	53,867	57,288
	#pairs	566	906	614	844

Notes: Data are from the SALT sample of the Swedish Twin Registry. This table reports the average yearly income between age 30 and 50, in SEK, for subjects, disaggregated by birth cohort, sex and zygosity. Income is defined as the sum of income earned from wage labor, income from own business, pension income and unemployment compensation. The sample is restricted to those individuals for whom there is no missing data (approximately 94% of the sample) and whose reported average yearly income exceeded 1000 SEK (approximately \$150).

**Table A2: Summary Statistics for Net Wealth**

	Men		Women	
	<u>MZ</u>	<u>DZ</u>	<u>MZ</u>	<u>DZ</u>
Net Wealth in SEK	2,400,000	3,550,000	1,990,000	1,500,000
S.D.	(6,150,000)	(30,700,000)	(10,100,000)	(6,740,000)
Standard Normal	0.28	0.16	-0.05	-0.12
S.D.	(1.03)	(1.06)	(1.00)	(0.91)
Hyperbolic Sine	13.26	12.69	11.77	12.07
S.D.	(5.99)	(6.72)	(7.61)	(6.74)
#pairs	175	174	213	189

Notes: Data are from the first wave of the SALTY sample of the Swedish Twin Registry. This table reports summary statistics for net wealth, disaggregated by zygosity and gender. Net wealth is defined as the difference between the total self-reported value of assets and total self-reported debt. The asset classes considered are: property (including summer house), stocks, bonds, transportation vehicles and other. Respondents are asked to prorate in cases of joint ownership of an asset or joint debt.

# Online Appendix II

## Sample Description

**Sample:** The Icelandic Heart Association (IHA) initiated the Reykjavik Study in 1967 as a random sample of 30,795 men and women born between 1907 and 1935 who were living in Reykjavik. Each participant was randomly assigned to one of six groups and, depending on the group, surveyed up to seven more times between 1967 and 1991. In 2002, the NIA Laboratory of Epidemiology, Demography, and Biometry in collaboration with IHA initiated the Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-RS) to collect genotypic as well as additional medical, cognitive, and survey data from 5,764 of the surviving participants. The AGES-RS sample was randomly selected from among the surviving Reykjavik Study participants, except that individuals who had been surveyed more than once were oversampled. Examinations of the participants occurred between 2002 and 2006. Participants underwent an extensive physical and physiological exam, including blood samples, as well as a survey based on the Reykjavik Study questionnaire (see Harris et al., 2007, for more details).

## Variable Description

Although designed to examine risk factors in relation to a range of diseases and disabilities in old age, AGES-RS collected a variety of data that we used to construct 8 outcome measures of interest to economists, which we describe here. Unless otherwise noted, the survey and exam data we use are from the AGES questionnaire, administered between 2002 and 2006 when respondents were 67 years old or older. In some cases, noted below, we use survey data from the Reykjavik Study, which were asked between 1967 and 1991, mostly in 1967.

**Human capital index.** We constructed the human capital index from a participant's responses to two questions:

What is the highest level or year of school that you completed?

- \_ Did not go to school
- \_ Elementary school
- \_ High school
- \_ Industrial College, midwife, nurse's aid, art/music education
- \_ Farmer's College
- \_ House keeping
- \_ Seamanship
- \_ Junior College
- \_ Business school
- \_ Teacher's College/Nurseing school
- \_ University/Technical College

To convert these response to a years of schooling measure, we consulted several historians of the Icelandic educational system. Jon Torfi Jonasson, then-head of the education department at the University of Iceland and editor of a history of the Icelandic school system, provided an initial set of estimates of average years completed for each of the above response categories in the years 1925 and 1955, adjusted for the individual's response to a survey question regarding whether they grew up in an urban area. He told us that he provided us with urban estimates that are closer to the official mandates, while the rural number attempts to adjust for the facts that (1) the mandates regarding elementary school were not strictly enforced in rural areas; and (2) widespread home schooling in rural areas compensated for some of the lack of formal schooling.

We next consulted Icelandic historians Loftur Guttormsson and Helgi Skúli Kjartansson. They made slight adjustments to Jonasson's estimates. For a respondent who completed schooling in 1925 or 1955, the estimates for each of the above response categories, in the form (1925 urban, 1925 rural, 1955 urban, 1955 rural) are, respectively:

(0, 0, 0, 0)  
(5, 3, 8, 4)  
(8, 4, 9, 7)  
(7, 5, 11, 8)  
(6, 4, 9, 6)  
(6, 4, 9, 5)  
(7, 5, 12, 7)  
(11, 9, 13, 11)  
(7, 5, 12, 7)  
(8, 6, 13, 11)  
(15, 13, 18, 16)



They also advised us that, due to two changes in schooling laws that affected urban students exclusively, we should add a “whole year of schooling to the cohort of 1929 above the one of 1928” and correct for the fact that a child “born in 1934 would spend an extra year at school compared with the one born in 1933.”

We categorize respondents as urban or rural using the survey question described under “Housing wealth” below. For non-urban respondents, for each response category, we modeled years of schooling for a respondent who completed schooling in year  $t$  as a linear function of  $t$  that passes through the estimates for 1925 and 1955. For urban respondents, for each response category, we similarly modeled years of schooling as a linear function, except with the two discontinuous jumps described above. We used the implied value of years of schooling completed as our estimate.

The second question used in the human capital index is “How many foreign languages have you learned to speak or read during your lifetime?” Respondents provided a numerical response. We dropped one outlier respondent whose answer was 35.

We z-scored and summed these two component variables to form the human capital index.

**Cognitive function index.** Cognitive function was *not* one of the 8 outcomes we initially studied. We examined it subsequently in order to better understand the association we found with years of education.

Scores on cognitive tests were assessed during an in-person exam. We created a cognitive function index by summing z-scores of performance on four cognitive tests: digit symbol substitution (a measure of information processing speed taken from the WAIS intelligence test), digit span (forward—a measure of working memory capacity, and backward—a measure of cognitive flexibility), spatial working memory, and long-term verbal memory (the CVLT memory test’s recall and recognition subtests). To restrict our focus to cognitive ability in the normal range of variation, we excluded participants who scored  $\leq 23$  on the Mini Mental State Exam, a standard cutoff for probable dementia. Finally, we z-scored the resulting variable so that the index has zero mean and unit variance.

**Income (as predicted from occupation).** Respondents answered the free-response question:

What type of occupation did you work at for the longest period of your working life and for how many years?

A research assistant, Ólafur Garðar Halldórsson, coded these responses into 4-digit occupational codes developed by Statistics Iceland. Using 2007 Statistics Iceland data, we then assigned respondents the mean income from their 4-digit occupational code. In some cases, we also used the response to the Yes/No question, “Are you a manager or a foreman?” (asked in the Reykjavik Study survey) to assign the appropriate mean income.

**Housing wealth.** We determined whether a respondent was a homeowner or not from the response to this question asked in the Reykjavik Study:

Are you:

- ☐ Owner of your apartment?
- ☐ Tenant?

We assigned tenants a housing wealth of zero.

For homeowners, we coded characteristics of the house from the Reykjavik Study questions:

What is the size of your apartment?

- ☐ 1 room
- ☐ 1 room and kitchen
- ☐ 2 rooms and kitchen
- ☐ 3 rooms and kitchen
- ☐ 4 rooms and kitchen
- ☐ 5 or more rooms and kitchen

How old is the house you live in?

- ☐ 0 - 9 years
- ☐ 10 - 19 years
- ☐ 20 years or more

Do you live in:

- ☐ A villa

- \_ A duplex
- \_ An apartment block

From the first question, we coded the number of rooms based on the responses as 1, 2, 3, 4, 5, or 6, respectively. From the second question, we coded the age of the house. From the third question, we coded the type of the house as single detached, single attached, and multiplex, respectively.

To help determine home value, we also used the response to this question from AGES:

Where did you live when you were between 40-59?

- \_ On a farm
- \_ In a fishing village
- \_ In a village
- \_ In the city

We created a dummy for urban status that equaled one for the response “In the city” and zero otherwise.

To translate the home characteristics to an estimate of house value, we estimated a hedonic regression using 2000 U.S. Census data. From the Census data, we dropped homes described as “Mobile home or trailer” or “Boat, tent, van, other,” as well as homes that are also used commercially. We used as housing value the respondent's estimate of the full current market value of the property, including both house and land. If a house had a number of rooms exceeding 6, we recoded it as 6 rooms to match the truncation in the Icelandic data. Home value is top-coded at 999998 in the Census data, so we estimated tobit regressions with  $\log(\text{home value})$  as the dependent variable. Our regression included state fixed effects. Omitting the fixed effects, the fitted equation is:

$$\begin{aligned} \text{Log}(\text{home value}) = & 0.226583 \times (\text{number of rooms}) - 0.2090792 \times (\text{age of house is 10-19 years}) \\ & - 0.4701176 \times (\text{age of house is 20 or more years}) - 0.1963042 \times (\text{single detached}) \\ & + 0.2533242 \times (\text{multiplex}) + 0.2748653 \times (\text{urban}) + 10.95979. \end{aligned}$$

We used the fitted value of this equation for the Icelandic respondents.

**Labor supply.** To construct a measure of labor supply for respondents who were employed, we used the responses to these two questions from the Reykjavik Study:

How many hours do you usually work each week?

1. Main occupation

\_ 0 - 29 hours

\_ 30 - 39 hours

\_ 40 - 49 hours

\_ 50 - hours

2 . Extra work.

\_ 0 - 9 hours

\_ 10 - 19 hours

\_ 20 - hours

For the first question, we coded responses as 15, 35, 45, and 55 hours, respectively. For the second question, we coded responses as 5, 15, and 25 hours, respectively, or 0 if missing. We then summed the number of hours, took the log, and standardized the resulting variable to have zero mean and unit variance.

**Happiness.** Our happiness measure is created from responses to the following Yes/No questions:

Are you basically satisfied with your life based on the feelings over the past week?

Do you feel that your life is empty based on the feelings over the past week?

Are you in good spirit most of the time?

Do you feel happy most of the time?

Scoring “Yes” as 1 and “No” as 0, we added the responses to the first, third, and fourth question and subtracted the response to the second question.

**Overall health.** We constructed an index of overall health from the responses to two questions from the Reykjavik Study:

How many days have you been absent from work because of illness in the last 12 months?

\_ None

\_ 1-7 days

- ☐ 8-14 days
- ☐ 15-21 days
- ☐ 22-28 days
- ☐ 29- days

In general, how would you say your health is? Would you say it is...

- ☐ Excellent
- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Poor

We standardized responses to each question to have mean zero and unit variance and summed these scores.

**Time discounting index.** Our time discounting capital index is composed of three subindices: drinking, smoking, and body mass index (BMI).

*Drinking.* These questions measure how often the respondent drank in midlife:

When you were aged 40-59 (in midlife) did you drink alcoholic beverages?

- ☐ Yes
- ☐ No

[If responded “Yes”:]

How often did you drink alcoholic beverages then?

- ☐ Daily
- ☐ 2 to 3 times a week
- ☐ 1 time a week
- ☐ 2 or 3 times a month
- ☐ 1 time a month
- ☐ Less than 1 time per month
- ☐ Don’t Know

We dropped the “Don’t know” responses. We coded “No” as 0, “Less than 1 time per month” as 1, “1 time a month” as 2, and so on, up to “Daily” as 7. Then we standardized the resulting variable to have zero mean and unit variance.

The next question measured the quantity the respondent drank:

On occasions when you drank, how much did you drink?

- ☐ One drink
- ☐ Two drinks
- ☐ Three drinks
- ☐ Four or more drinks
- ☐ Don't Know

We dropped the “Don’t know” responses, and we coded non-drinkers as 0. We coded the other responses as 1, 2, 3, and 4, respectively. Then we standardized the resulting variable to have zero mean and unit variance.

These questions measure the length of time the respondent drank alcohol:

Was there another time in your life that you did drink alcoholic beverages?

- ☐ Yes
- ☐ No
- ☐ Don't Know

[If responded “Yes”:] At what ages did you drink alcoholic beverages then?

- ☐ From (age)
- ☐ To (age)

We took the difference between the “To” and “From” responses to get the number of years the respondent drank alcoholic beverages. We set the number of years equal to 0 if the respondent indicated never having drunk alcoholic beverages. We set it equal to 5 if the difference was negative because such a respondent had clearly drunk alcoholic beverages at some point but probably for a short time, and 5 was on the far low end of the distribution of number of years. We took  $\log(1 + \text{number of years})$  and then standardized the resulting variable to have zero mean and unit variance.

This Yes/No question is a measure of whether alcohol consumption was ever excessive:

Has someone close to you ever told you that your alcohol use was having a negative effect on your home, work or relationships?

- ☐ Yes
- ☐ No
- ☐ Don't Know

We dropped the “Don’t Know” responses, coded “No” as 0 and “Yes” as 1, and then we standardized the resulting variable to have zero mean and unit variance. The alcohol subindex is the sum of these four components.

*Smoking.* There are two Yes/No questions regarding whether the respondent currently smokes cigarettes or ever smoked cigarettes, and a question asked of any respondent who answered yes to either question:

Do you smoke cigarettes now?

If you do not smoke cigarettes now, did you ever smoke cigarettes regularly?

About how old were you when you first started smoking cigarettes (years of age)?

Any respondent who indicated having smoked at some point but not smoking currently was also asked:

About how old were you when you quit smoking cigarettes (years of age)? If you are not sure, give your best guess.

We calculate the number of years the respondent has smoked cigarettes using responses to these questions. We set the number of years equal to zero for respondents who said they do not smoke now and have not ever smoked. For any respondent who smokes now or has smoked and has not quit, we set the number of years equal to the current age minus the age when the respondent started smoking. For any respondent who quit smoking, we set the number of years equal to the age when the respondent quit minus the age when the respondent started smoking. For any respondent where this calculation yields a negative estimate of number of years smoked, we set the number of years equal to 5 because such a respondent had clearly smoked at some point but probably for a short time, and 5 was on the far low end of the distribution of number of years. We took  $\log(1 + \text{number of years smoked cigarettes})$  and standardized the resulting variable to have zero mean and unit variance.

To get an estimate of the quantity of cigarettes smoked, we used responses to these questions:

On average, about how many cigarettes per day do you usually smoke? (Less than 1 cigarette per day?)

When you were smoking, about how many cigarettes did you smoke (per day)? If you are not sure please make your best guess. Less than 1 cigarette per day?

What is the most cigarettes per day you have ever smoked for as long as a year? If you are not sure, please make your best guess.

We coded any response of less than 1 per day as 0.5 per day. For a respondent who currently smokes, we estimated the amount of cigarettes per day by taking an average of the number per day currently smoked and the maximum number per day smoked. For a respondent who currently does not smoke but did smoke at some point, we use the number of cigarettes per day the respondent said he or she did smoke. For a respondent who currently does not smoke and never smoked, we set the number per day to be zero. We took  $\log(1 + \text{number of cigarettes smoked per day})$  and standardized the resulting variable to have zero mean and unit variance.

Our overall smoking subindex is the sum of the measure of number of years smoked and the measure of number of cigarettes smoked per day.

*BMI.* For the BMI subindex, the key survey question is:

What was the most you have ever weighed (in kilograms) (women: answer for a time when you were not pregnant)?

Height (in meters) was measured during an in-person exam. We calculated BMI as  $(\text{max. weight in kg}) / (\text{height in cm})^2$ . We took the log of BMI and then standardized the resulting variable to have zero mean and unit variance.

Our time discounting capital index is the sum of the three subindices.

**Social capital index.** Our social capital index is composed of two subindices, one for family and one for non-family. Our family index aggregates information from three questions:

Other than those who live with you, how often do you meet with your children or relatives?

- ☐ Daily
- ☐ Weekly
- ☐ Monthly
- ☐ Yearly
- ☐ Never
- ☐ I have no living children or living relatives



How often are you in telephone contact with your children or relatives?

- ☐ Daily
- ☐ Weekly
- ☐ Monthly
- ☐ Yearly
- ☐ Never
- ☐ I have no living children or living relatives

If you have living relatives, about how many relatives do you feel close to, feel at ease with, can talk to about private matters, and call on for help?

- ☐ I have no living relatives

The responses to the first question were coded as 4, 3, 2, 1, 0, and 0, respectively, and then standardized to have zero mean and unit variance; the second question was coded identically.

For the third question, we took  $\log(1 + \text{number of relatives})$  and then standardized the resulting variable to have zero mean and unit variance. The family subindex is the sum of these three variables.

Our non-family index aggregates information from six questions:

Other than those who live with you, how often do you meet with your friends or neighbors?

- ☐ Daily
- ☐ Weekly
- ☐ Monthly
- ☐ Yearly
- ☐ Never

How often are you in telephone contact with your friends or neighbors?

- ☐ Daily
- ☐ Weekly
- ☐ Monthly
- ☐ Yearly
- ☐ Never

[Asked immediately after the question above about living relatives you feel close to:]

About how many others do you feel close to, feel at ease with, can talk to about private matters, and call on for help?

In the past 12 months, how often did you participate in church, community, or social club activities (in addition to those mentioned above)?

- ☐ Not at all

- ☐ Less than once a month
- ☐ At least once a month
- ☐ At least once a week
- ☐ Daily

How often do you usually attend religious services, meetings or activities?

- ☐ Never
- ☐ A few times a year
- ☐ 1 - 3 times per month
- ☐ Once per week
- ☐ 2 - 3 times per week
- ☐ Nearly every day

[Asked immediately after the question above about non-relatives you feel close to:]  
About how many of them do you see or hear from at least once a month?

The responses to the first three question were coded the same as the analogous questions regarding family members. The responses to the fourth and fifth questions were coded as 1-5 and 1-6, respectively, and then standardized. For the sixth question, we took  $\log(1 + \text{number of relatives})$  and then standardized the resulting variable to have zero mean and unit variance. The non-family subindex is the sum of these six variables.

The social capital index is the sum of the two subindices.

## References

Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, Sigurdsson G, Thorgeirsson G, Aspelund T, Garcia ME, Cotch MF, Hoffman HJ, Gudnason V. 2007. Age, Gene/Environment Susceptibility-Reykjavik Study: Multidisciplinary Applied Phenomics. *American Journal of Epidemiology*. 165(9): 1076–1087